# Structure-Function Prediction of Insect Odorant Binding Proteins

## Thesis

# Structure-Function Prediction of Insect Odorant Binding Proteins

A thesis submitted for the degree of Doctor of Philosophy

to

The Open University

by

George T. Tzotzos PhD

September 2012

Department of Life, Health and Chemical Sciences
Open University
Walton Hall
Milton Keynes MK7 6AA
UNITED KINGDOM

DATE OF SUBMISSION = 30 SEPTEMBER 2012
DATE OF AWARD = 22 JULY 2013

## Declaration

---

The work embodied in this thesis was carried out by the author between October 2005 and September 2012. This thesis was completed under the supervision of Dr Elaine A. Moore, Dr David R. Roberts, Professor James N. Iley and Dr Clare E. Sansom.

The work draws upon data found in the public domain; consequently the collection, analysis and processing of data has been carried out by none other than myself.

I declare that, except where stated, the work presented is the result of my own investigations. The material within this thesis has not been submitted, nor is currently being submitted, for any other degree.

A minor part of this works has been published/presented as listed below.

X. He, G. Tzotzos, C. Woodcock, J. A. Pickett, A. Hooper, L. M. Field & J-J Zhou. "Binding of the General Odorant Binding Protein of Bombyx mori BmorGOBP2 to the Moth Sex Pheromone Components". Journal of Chemical Ecology, 2010 DOI 10.1007/s10886-010-9870-7

Sixth International Symposium on Molecular Insect Sciences 1-5 Oct 2011 Amsterdam, The Netherland. Poster Presentation. X-L H1, L. M. Field, N. H. Keep, C. Cambillau, G. Tzotzos, J-J Zhou "Novel Structural Features of Insect Odorant Binding Proteins and Their Involvement in Ligand Binding"

## Acknowledgements

I would like to thank Dr Elaine A. Moore and Dr David R. Roberts for all their advice, support and encouragement and patience throughout my work. Their help has been appreciated more than they may realise. I also thank Professor James N. Iley for his support during the initial stages of my research and Dr Clare E. Sansom for introducing me to the Open University.

I am most thankful to my friend and colleague Dr Magnus Bosse for reducing my professional burden thus allowing me to dedicate precious time to my research.

Dedicated to my wife Susan Jane and my daughter Danae, for their encouragement and tolerance. Without their help I might have not completed this thesis.

## Abstract

This study concerns the application of bioinformatic tools for the elucidation of the biological function of insect general odorant and pheromone binding proteins (GOBPs / PBPs). These proteins are thought to function as transporters of volatile odorant molecules to olfactory receptors (ORs) situated in olfactory receptor neurons (ORNs) in insect antennae. Activation of ORNs by the odorant molecules gives rise to action potentials resulting in spatially defined patterns of glomerular activity in the brain, odour discrimination and concomitant behavioural response of the insect.

The extent to which OBPs are critical for olfactory discrimination remains unclear. Numerous hypotheses have been postulated regarding the ability of OBPs to discriminate specific odorants and/or pheromones as well as their playing a role in the activation of odorant-responsive chemosensory neurons, in functioning as selective filters in odour recognition or participating in signal termination by inactivating odorant molecules.

*In silico* binding studies of ligands and pheromones on OBPs derived from crystallographic studies or *de novo* homology modelling have been conducted primarily by docking and molecular dynamic (MD) simulations. It is shown that results obtained from such studies can provide useful insights and testable hypotheses with regard to the biochemical function of OBPs.

Docking and MD simulations corroborate experimental evidence that the *B. mori* general odorant binding protein (*Bmor*GOBP2) has considerably higher affinity than the *B. mori* pheromone binding protein (*Bmor*PBP1) for the pheromonal components bombykol and bombykal and predict that this is also true for the modelled *M. sexta* proteins (*Msex*GOBP2 and *Msex*PBP1). In addition, steered molecular dynamics (SMD) simulations predict ligand entry and exit pathways into and out of *Bmor*GOBP2.

In addition, docking and molecular dynamics (MD) simulations with the highly homologous odorant binding proteins from *A. gambiae* (*Agam*OBP1), *A. aegypti* (*Aaeg*OBP1) and *C. quinquefasciatus* (*Cqui*OBP1) provide evidence of differential capacity of these proteins to select ligands with specific structural characteristics.

## List of Tables

## List of Figures

## Abbreviations

| | |
|---|---|
| (G)OBP | (General) Odorant binding protein |
| PBP | Pheromone binding protein |
| OR | Olfactory receptor |
| OR(S)N | Olfactory receptor (sensory) neuron |
| ODE | odour degrading enzyme |
| AB(S)P | Antennal binding (specific) protein |
| SNMP | Sensory neuron membrane protein |
| IR | Ionotropic glutamate receptor |
| DEET | *N,N*-diethyl-*m*-toluamide |
| PEG | Polyethylene glycol |
| MOP | (*5R,6S*)-6-acetoxy-5-hexadecanolide |
| cVA | 11-*cis* vaccenyl acetate |
| BIS | Bisabolol |
| IUPAC | International Union of Pure and Applied Chemistry |
| MD | Molecular dynamics |
| SMD | Steered molecular dynamics |
| PME | Particle mesh Ewald |
| SASA | Solvent accessible surface area |
| MMPBSA | Molecular mechanics Poisson Boltzmann surface area |
| MMGBSA | Molecular mechanics Generalised Born surface area |

# SUPPLEMENTARY MATERIAL CD

## Docking Simulations

**ANNEX 1a Moth OBPs ligand library**

**ANNEX 1b Results of virtual screening (moth OBPs)**

**ANNEx 2a Mosquito OBPs ligand library**

**ANNEX 2b Results of virtual screening (mosquito OBPs)**

## Molecular Dynamics[1]

### Moth MD simulations

***BmorOBP1* - bombykol**

***BmorOBP1* - bombykal**

***BmorGOBP2* - bombykol**

***BmorGOBP2* - bombykal**

***MsexOBP1* - bombykol**

***MsexOBP1* - bombykal**

***MsexGOBP2* - bombykol**

***MsexGOBP2* - bombykal**

***SMD: BmorGOBP2* exit pathways**

### Mosquito MD simulations

***AgamOBP1-, AaegOBP1, CquiOBP1* - MOP**

***AgamOBP1-, AaegOBP1, CquiOBP1* - DEET**

***AgamOBP1-, AaegOBP1, CquiOBP1* - BIS**

***AgamOBP1* - indole**

---

[1] Each MD simulation contains (1) Antechamber (preparation of topologies); (2) restrained minimisation; (3) unrestrained minimisation; (4) heat equilibration; (5) density equilibration; (6) equilibration; (7) MD production; (8) MMPBSA; (9) Entropy; (10) per-residue decomposition of free energy; (11) Trajectory analysis

# Contents

*Chapter 1*

# *Insect Olfaction*

# 1.1 Introduction

Chemoreception is a crucial sensory modality for humans as well as other animals. It plays an essential role in almost every aspect of animal life. Animals possess highly sensitive and discriminating chemosensory systems which modulate foraging, mating, aggregation and flight behaviour. The detection of chemical cues in the environment is therefore essential for survival. Chemoreception is the combined term for the senses of taste (gustation) and smell (olfaction). The former requires contact of the sense organs with chemicals and is defined as contact-chemoreception, whereas the latter, at least in non-aquatic animals, is stimulated through contact with vapourised chemicals (distance-chemoreception). Understanding how animals detect literally thousands of odorant molecules by their chemosensory systems and how they subsequently process and translate this chemosensory information into a multitude of perceptions by their central nervous systems has been one of the major challenges of modern neurobiology. Linda Buck and Richard Axel were awarded the 2004 Nobel prize in Physiology and Medicine for their groundbreaking discovery of vertebrate olfactory receptors and contribution to the understanding of the underlying mechanisms of odour perception.

Both in higher animals and in insects, the recognition and discrimination of odorant compounds occurs in specialised sensory organs and is mediated by a large repertoire of receptors and signaling pathways. Olfactory receptor neurons (ORNs), also known as olfactory sensory neurons (OSNs) are located in peripheral systems and are responsible for odour detection, whereas odour discrimination takes place centrally in the brain. Activation of ORNs triggers action potentials resulting in spatially defined patterns of glomerular activity in the brain.

An understanding of how odours are encoded in the olfactory system is essential in deciphering how olfactory memory is stored in the brain.

Contrary to humans who experience the world mainly through visual and auditory systems, insects perceive the world through the detection of chemicals and as such have been described as "analytical chemists par excellence" [1]. The perception of chemical stimuli from the external world requires physical contact between foreign molecules and receptors situated in sensory neurons. Volatile organic molecules that modulate insect behaviour are known as semiochemicals. They are further subdivided into two main categories, namely, pheromones and allelochemicals, each of which consists of several subcategories. The division of semiochemicals into specific categories and subcategories is shown in Table 1.1. The recognition of pheromones is thought to be species-specific, whereas the recognition of allelochemicals is more broadly tuned.

| Semiochemicals | | | |
|---|---|---|---|
| **Pheromones** | | **Allelochemicals** | |
| **Type** | **Function** | **Type** | **Function** |
| Sex pheromones | Modulation of mating behaviour | Allomones | Conferment of benefits to emitter |
| Aggregation pheromones | Induction of insect aggregation | Kairomones | Conferment of benefits to receiver |
| Trail pheromones | Recruitment by workers of social insects of other individuals to food source or colony site | Synomones | Mediation of mutualistic beneficial interactions (emitter and receiver) |
| Alarm pheromones | Stimulation of flight or defense behaviour | | |
| Other | ..... | | |

**Table 1.1** **Categories and subcategories of semiochemicals**

Binding of semiochemicals on the receptor triggers the activation of signaling pathways and response to the chemical stimuli. Odour discrimination is thought to be achieved by the combined action of odorant binding proteins (OBPs) and

odorant receptors (ORs). OBPs provide a first filter of odorant discrimination by facilitating the solubilisation and transport of semiochemicals to ORs, whereas the latter provide a more specific filter of odour recognition. OBPs are the focus of this work. The term is used to describe both pheromone binding proteins (PBPs) as well as general odorant binding proteins (GOBPs).

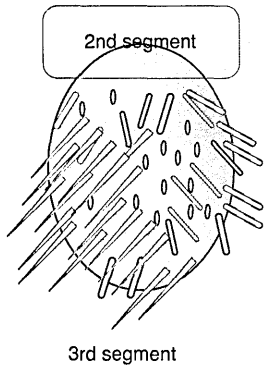
## 1.2 The 'anatomy' of the insect olfactory system

In insects, odours are detected by the ORNs located in two head appendages on the front of the head, the antennae, and maxillary palps. Antennae and maxillary palps are generally regarded as olfactory organs and tarsi and mouth as sensory organs. Antennae can process information related to motion and orientation, odour, sound, humidity, as well as chemical cues. There is considerable morphological variation in insect antennae but, in general, they consist of three segments. The first segment is known as the scape (base), the second as the pedicel (stem) and the third as the flagellum. The insect maxillary palp is a small antenna-like structure arising from the maxilla, a pair of mouthpart structures immediately posterior to the mandibles. These two organs bear cellular structures known as sensilla, which are formed of a cuticular wall that defines a cavity (the sensillar lymph) containing the dendrites of sensillar neurons. The lymph is porous allowing passage of volatile chemicals and prevents evaporation of the fluid filling the lymph.

The antennae are covered by three types of sensilla, namely, basiconic, trichoid and coeloconic and these are further subdivided into distinct morphological subclasses [2]. The maxillary palp, on the other hand, is covered by approximately

60 basiconic sensilla. Each sub-class of sensilla responds to different odour types (Figure 1.1).

Each sensillum is surrounded by support cells, which arise from mitotic activities from one or a small group of defined precursor cells during sensillogenesis. Three types of support cells are known, namely, trichogen, tormogen and tecogen [3].




| | Sensilla types | Stimuli |
|---|---|---|
| | basiconic | Carbon dioxide<br>General odorants |
| | coeloconic | General odorants<br>Small amines<br>Humidity |
| | trichoid | Pheromones |

**Figure 1.1** **Cartoon of antenna, types of chemosensory hairs and classes of odorant stimuli**

In Drosophila, sensilla are innervated by 1-4 ORNs which project their axons along the antennal nerve to the antennal lobe glomeruli where they are sorted according to chemosensitivity. In other insects, sensilla may be innervated by as many as 30 ORNs. From the glomeruli, information is relayed by projection neurons in the inner and medial antennocerebral tract to the mushroom body and to the lateral horn [4].

For olfaction to occur, three conditions are necessary, namely, (a) passage and solubilisation of the foreign volatile molecules into the sensillar lymph, (b) a mechanism to prevent evaporation of water and protect the sensory neuron endings from the environment and (c) physical contact between foreign molecules and receptors of the sensory neurons. Physical contact between the solubilised

foreign molecules and the olfactory receptors (ORs) located on the ORNs is mediated by soluble polypeptides which are found in the 'perireceptor space' and act as transporter molecules. The latter are commonly known as odorant binding proteins (OBPs). The biochemistry of odour detection involves yet another type of protein, namely, the odour degrading enzymes (ODEs). The chemical interactions between the solubilised foreign molecules (e.g. odours, pheromones), the OBPs and ODEs are referred to as 'perireceptor events' [5]. Receptor events, that is interactions between the odorant molecules and the ORs, are then translated into an ion gradient potential signal by transductory proteins and/or ion channels which is recognised by the central nervous system. A simplified picture of perireceptor event is depicted in Figures 1.2.a and 1.2.b below.

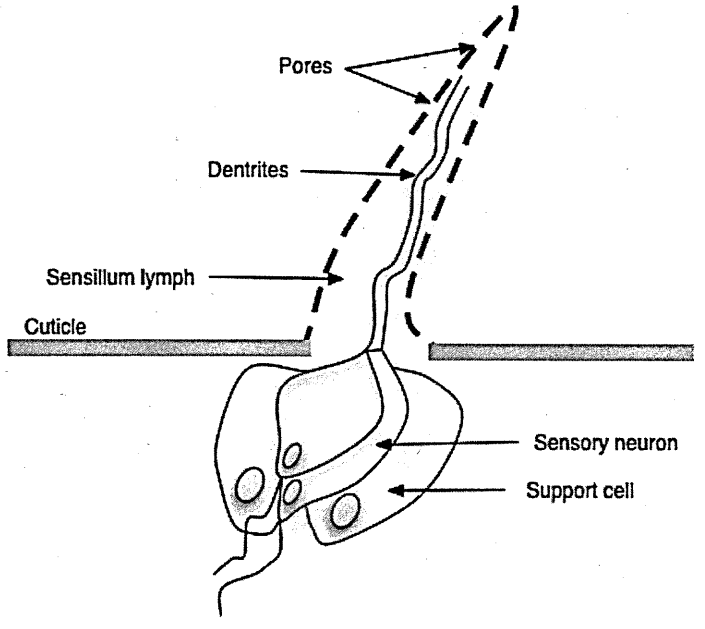



**Figure 1.2.a** **General structure of an insect olfactory hair**
Figure modified from J. Rozas *et al.* [54]

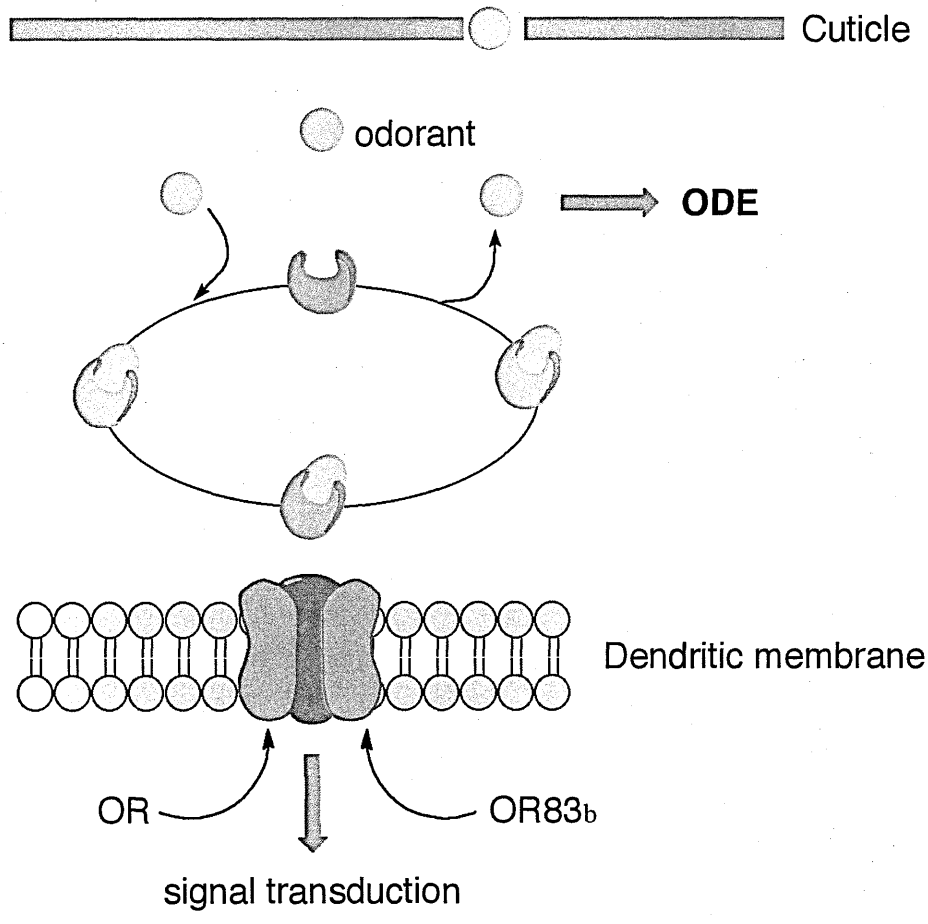



**Figure 1.2.b Perireceptor events of the chemosensory transduction pathway**
A simplified functional scheme of perireceptor events taking place at the dendrites of ORNs (see 1.2.a). An alternative scheme is shown in Figure 1.5. Scheme modified from J. Rozas *et al.* [54]

# 1.3 The biochemistry of insect olfaction

## 1.3.1 Perireceptor events: the role of OBPs

Perireceptor events take place in the lymph of the olfactory sensilla which provides an aqueous environment rich in fatty acids and odorant binding proteins (OBPs). Fatty acids provide the medium for the solubilisation of volatile, primarily hydrophobic odorant molecules. The concentration of these carrier proteins in the sensillum lymph can be as high as ~10 mM [5].

OBPs are part of a larger family of transport proteins, members of which are also expressed in non-olfactory tissues suggesting that these proteins may function as

broad specificity carriers for lipophilic compounds. The term 'encapsulins' has been coined for this larger protein family [6]. OBPs have been grouped into three sub-families, namely, pheromone binding proteins (PBPs), general odorant binding proteins (GOBPs) and antennal binding proteins (ABPs). The classification of OBPs into distinct sub-families is not without pitfalls and is based on a number of criteria including a unique signature of six cysteine residues forming three disulfide bonds. PBPs are expressed in auxillary cells and are secreted into the lumen where the mature protein can be detected in the sensillar lymph. They occur predominantly in the sensillar lymph of the male antennae although female PBPs are also known. In some cases, PBPs have been found in male sensilla that are not pheromone-sensitive [7], [8]. It appears likely that they may recognise general odorants as well. GOBPs are expressed in the antennae of both male and female insects. They are assumed to be involved in the detection and transport of semiochemicals other than pheromones. ABPs show little sequence similarity to either GOBPs or PBPs but have the same six cysteine signature [9].

Experimental evidence to date is inconclusive as to the precise function of OBPs. Several hypotheses have been put forward which are not necessarily mutually exclusive and may be system specific. These are summarised below.

***OBPs/PBPs are non-essential for ligand specificity.*** Experiments involving seven sex-pheromones from three moth species and their respective olfactory receptors expressed in *Xenopus* oocytes in the absence of PBPs responded dose-dependently only to the main sex-pheromone of the corresponding moth species [10]. The results of these experiments were in line with observations involving *A. gambiae* olfactory receptors which were shown to respond with high specificity to ligands in the absence of PBPs [11].

***OBPs/PBPs are required for the transport of hydrophobic odorants.*** Several studies have demonstrated that OBPs are involved in the recognition and solubilisation of airborne hydrophobic odorants and pheromones and subsequent delivery to the olfactory receptors [12] [13] [14].

***OBPs/PBPs may act as ligand protectors or scavengers of excess ligand.*** Odorant molecules are liable to degradation by odorant degrading enzymes (ODEs). It has been suggested that the entrapment of odorant in the binding pockets of OBPs may protect them from enzymatic degradation by lymph ODEs. Alternatively, OBPs may "scavenge" excess ligand to avoid secondary stimulation of the neurons. This mechanism would be an alternative to ligand degradation by ODEs [15] (see Section 1.3.2). The PBP of the gypsy moth, *Lymantria dispar*, binds its natural pheromone (*7R,8S*)-2-methyl-7,8-epoxyoctadecane, with some discrimination which increases in the presence of its known antagonist, (*7Z*)-2-methyloctadec-7-ene. Binding enhancement also occurs at high ligand:PBP ratios, suggesting that the PBP may function both as a pheromone transporter and as scavenger [16].

***OBPs/PBPs may be essential for signal transduction.*** Experiments with *D. melanogaster* T1 neurons sensitive to the aggregation pheromone (*Z*)-11-octadecenyl-1-acetate, also known as *cis*-vaccenyl acetate (cVA) showed that the OBP LUSH played a specific role as a signal transduction component. Cells lacking LUSH were completely devoid of evoked activity to cVA. Transgenic expression of LUSH restored behavioural response [17].

***OBPs/PBPs may be enhancers of olfactory sensitivity.*** The sensitivity of T1 neurons to cVA increases over 500-fold in the presence of LUSH as compared to cells devoid of LUSH [18]. Similarly, experiments involving the expression of the *B.*

*mori* OR1 (*BmorOR1*) in an "empty neuron" housed in ab3 sensilla of *D. melanogaster*, showed significantly higher electrophysiological responses to the pheromone bombykol in the presence of the *B. mori* PBP as compared to those devoid of it [19]. In the same experiments, it was shown that bombykol can stimulate directly fly cells expressing *BmorOR1* but lacking the *B. mori* PBP suggesting that it is the pheromone alone and not the PBP-pheromone complex that activates the OR.

***OBPs/PBPs have broad ligand specificity.*** The selectivity and discriminating capacity of OBPs has been tested by means of binding assays to specific semiochemicals and pheromones. The results obtained from these studies indicate broad specificity binding not only for OBPs but also for PBPs. A number of such examples are presented in Section 1.3.3.

***OBPs/PBPs may function as selectivity filters.*** Selective binding of odorants and/or pheromones to different OBPs has also been shown. For example, the antennal-specific protein (ASP1) of the honeybee shows high specificity binding to the major components of the queen pheromone [20]. Very selective binding has also been demonstrated with PBPs of the gypsy moth. Two pheromone binding proteins found in the *L. dispar* antennae were shown to differentiate the (+) and (-) enantiomers of the sex pheromone disparlure [21]. In other studies, involving receptor cells of the sensilla trichodea of the silkmoth *Antherea polyphemus*, it was shown that three different types of PBPs, in various combinations with pheromone components, elicited electrophysiological impulses in receptor cell types that are not activated by the PBPs under physiological conditions, suggesting that both the pheromone components and the PBP are required for receptor activation [22]. It is therefore possible that the binding of specific pheromones to PBPs may activate

the latter through the induction of conformational changes. This has been demonstrated in elegant structural studies involving the *D. melanogaster* PBP, LUSH, in complex with the aggregation pheromone *cis*-vaccenyl acetate (cVA). It was shown that both cVA and LUSH are required for receptor response. It was also shown that cVA induces a major conformational change at the C-terminal of LUSH. A LUSH mutant locked in the active conformation by mutating Asp118 to Ala induced electrophysiological response in T1 sensilla in the absence of cVA [18].

Such conformational changes of a OBP-odorant complex can be induced by pH changes as shown in the case of the *B. mori* PBP1 (*Bmor*PBP1) in complex with the natural pheromone bombykol. The pheromone is transported by *Bmor*PBP1 through the sensillar lymph until it reaches certain negatively-charged sites on the dendrites. There, the low pH microenvironment induces a conformational change of the *Bmor*PBP1-pheromone complex with concomitant release of bombykol [23]. Following stimulation of the receptor, bombykol is inactivated.

The above observations point to the conclusion that OBPs can selectively transport pheromones and other semiochemicals to their respective receptors. It has been proposed that the overall selectivity of the OBP/OR system is achieved by "layers of filters" [6]. According to this proposition, OBPs transport only a subset of odorant compounds that enter the sensillum lymph, some of which might not bind to the receptors housed in the particular sensilla. Likewise, ORs have been shown to be broadly tuned to different compounds [24]. Olfactory response is triggered only by the fraction of compounds that are selected by both ORs and OBPs and thus enhanced specificity can be achieved if OBPs and ORs function in a two-step filter.

### 1.3.2 Perireceptor events: the role of ODEs

The insect olfactory system, in addition to its ability to discriminate odorants in minute amounts, requires another process by means of which signals generated upon exposure to a particular stimulus are inactivated. Insect flight through an odorant plume requires rapid behavioural responses to discontinuous changes in ambient concentrations of semiochemicals. These changes are reflected by biochemical reactions in the perireceptor space. Signal detection and termination at the ORNs occurs on the millisecond scale [6]. The most likely process to achieve signal termination is odorant inactivation either by degradation or other means (e.g. sequestration). Odorant-degrading enzymes (ODEs) found in the sensillum are thought to have been evolved precisely for this purpose [25]. It is reasonable to expect that different ODEs would attack specific functional groups such as esters, aldehydes, ketones, alcohols and epoxides and as such could be potential biotechnological targets for insect control as their inhibition would result in odour misperception. Although the first such pheromone degrading enzyme (*Apol*PDE), a sensillum esterase from *A. polyphemous*, was identified as early as 1981 [26] little is known about their actual function probably due to the fact that the concentration of ODEs is extremely low. *Apol*PDE expressed in the antennae of *A. polyphemus*, for example, was found to be ~20,000 lower than that of *Apol*PBP. The efficiency of this enzyme in catalysing the inactivation of excess pheromone was found to be consistent with the temporal resolution for odorant-mediated flight through a pheromone plume [27].

Any hypothesis on the mode of action of ODEs has to address the question of how it is possible that semiochemicals are not degraded by ODEs prior to their interaction with the ORs. Two likely models have been proposed to account for the mode of action of ODEs based on experimental and modelling data with moth

PBPs [28]. The first model, 'carrier-to-scavenger model' proposes that the PBP acts as a transporter of the pheromone through the extracellular sensillum lymph and mediates its interaction with the OR. Subsequently, through the action of a postulated enzyme, the PBP undergoes a conformational change turning into a scavenger of the pheromone. According to the second model, 'scavenger model', the free pheromone interacts with the OR without the mediation of the PBP, which acts as a scavenger only. The latter model is an unlikely one as it would require that odorant molecule contact the ORs directly. The possibility could arise if semiochemicals interacted with ORs directly via tubule-dendrite contacts. However, this does not seem to be a likely route [29].

### 1.3.3 Receptor events: the role of ORs

ORs were first discovered in 1991 in the rat olfactory epithelium [30]. The discovery was based on the correct assumption that ORs would be encoded by genes expressed only in olfactory tissues and be members of the G-protein coupled receptor (GPCR) superfamily as the latter are known to be involved in a variety of cellular processes such as signal transduction, neurotransmission and hormonal regulation. Genomic searches for ORs were based on homology approaches for the identification of conserved transmembrane domains of the GPCR family, annealing of the conserved regions with degenerate oligonucleotide primers and amplification and identification of complementary DNAs (cDNAs) of OR encoding genes by means of the polymerase chain reaction. Since that time, OR genes have been identified in numerous vertebrates including humans as well as in the nematode *C. elegans*.

G-protein coupled receptors are proteins comprising seven α-helix transmembrane domains which form complexes with a heterotrimeric transducer (G protein) and

one of several effector enzymes such as phospholipase or adenyl cyclase that are involved in the synthesis of secondary messengers such as inositol 1,4,5-triphosphate or cyclic AMP. The N-terminus of the ORs is located extracellularly whereas the C-terminus is found within the cell cytoplasm. The transmembrane helices are linked with loops, three of which are extracellular and three intracellular. This topology has been confirmed through the crystallographic structure determination of bovine rhodopsin which remains the only GPCR structural model solved to date [31].

Candidate insect ORs were identified initially during the genome sequencing project of *D. melanogaster* by research groups which used computational algorithms to search for exons that encode seven helix transmembrane proteins [32] [33]. The genomic sequencing of several insect species has led to the identification of their respective ORs repertoires through a combination of bioinformatic and functional studies. These proteins have been shown to adopt a reverse topology to that of mammalian ORs, that is the N-terminal domain of the protein is found in the cytoplasm and the C-terminal domain is located extracellularly [34] [35] [36].

A sequence comparison of ORs has shown that sequence identity between ORs is low, even among species of the same insect order, with one notable exception, namely that of the Drosophila OR83b. This receptor is expressed in all but one type of sensory neuron. Orthologues of this protein, sharing sequence identity of approximately 60-80%, have been found not only amongst the different Drosophila species but also in species of other insect orders. It has been shown that OR83b alone is not involved in odour recognition [37]. However, in most Drosophila olfactory neurons ORs are co-expressed with OR83b. Furthermore, it has been

demonstrated that in mammalian cell lines or *Xenopus laevis* oocytes expressing insect ORs, the cellular response to semiochemicals increases dramatically if a given OR is co-expressed with OR83b or its orthologues. These findings suggests that oligomerisation of ORs with OR83b may be a necessary condition for initiating signal transduction and that OR83b may function as a co-receptor of the tuning OR [38]. Another hypothesis suggests that OR83b may also function as chaperone assisting in receptor trafficking and targeting. In mutants lacking OR83b, dendritic localisation of insect ORs is abolished and cellular response to odorants is lost [39].

More recently, two independent studies led to the conclusion that insect odorant signaling pathways are mediated by ligand-gated ion channels and that ORs are ionotropic receptors. However, on the basis of experimental evidence, the mechanisms that were proposed differ significantly. In one of the studies it was found that odour induced depolarisation is mediated by the opening of an ion channel formed between a common receptor (ORX) and the OR83b co-receptor [40] (Figure 1.3).

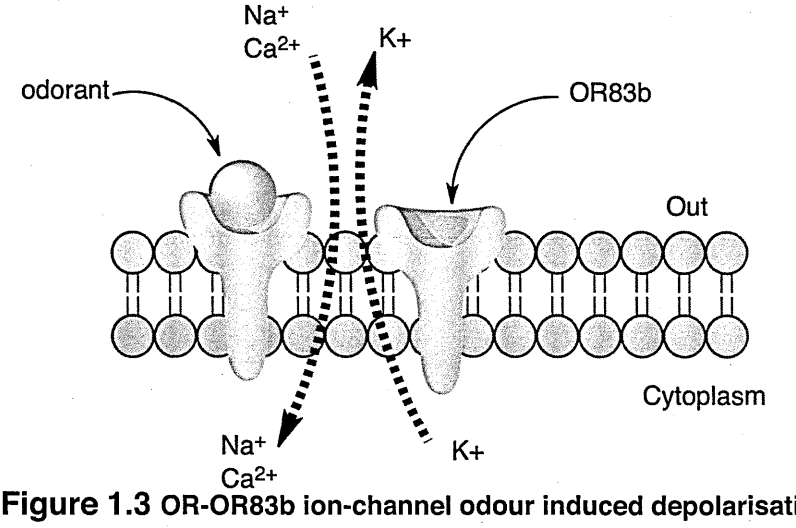



**Figure 1.3 OR-OR83b ion-channel odour induced depolarisation**

Modified from U. B. Kaup [210]

The second study proposes a dual mechanism involving an ion channel as well as the involvement of G proteins and/or intracellular messengers such as cAMP, cGMP or inositol-1,4,5-triphosphate [41] (Figure 1.4). The two studies leave a number of open questions concerning whether OR83b is a co-determinant of ligand affinity and selectivity, the exact stoichiometry of the heteromer formed and how signal transduction is terminated.

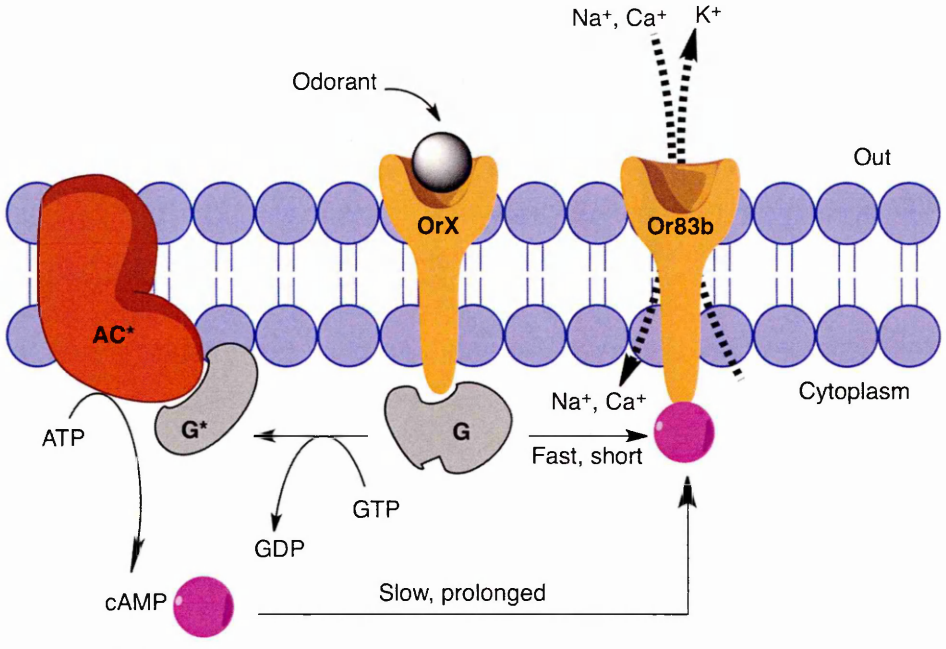



**Figure 1.4** **Alternative hypothesis of olfactory signal transduction**
Modified from U. B. Kaup [210]. AC stands for adenylate cyclase

Pheromone receptors, which are a sub-set of the OR superfamily, appear to require for proper function a sensory neuron membrane protein (SNMP) in addition to OR83b. The involvement of SNMP in olfactory signal transaction has been documented in the case of *D. melanogaster*. OR67d has been associated with the recognition of *cis*-vaccenyl acetate (cVA) a known *D. melanogaster* pheromone. It was shown in *in vitro* studies that the heteromeric complex OR67d:OR83b:SNMP is activated by high concentrations of *cis*-vaccenyl acetate (cVA). Thus, it seems that the activation of this receptor requires the presence of both OR83b and SNMP [42] [43]. However, *in vivo*, complex activation required also the presence

of LUSH [17]. The latter is an OBP formed in the lymph of a sub-set of trichoid ORNs. Mutants lacking the gene that encodes LUSH do not respond to cVA.

Another family of protein receptors has been implicated in olfaction and gustation, namely the family of ionotropic glutamate receptors (IRs). Phylogenetic studies have revealed two sub-families of IRs: a conserved "antennal IR" sub-family which is implicated in olfaction and a species-specific sub-family of "divergent IRs" which is expressed in peripheral and internal gustatory neurons. There is little available data on IR sensory neuron ligand interactions but the few that are available to date suggest that IRs are responsive to carboxylic acids, water, ammonia and other small molecules which are known to elicit broad behavioural responses to many insect species. ORs on the other hand may be tuned to detecting species-specific odour cues. Two IRs (IR8a and IR25a) are ubiquitously expressed in coeloconic sensilla. It is hypothesised that they may function as co-receptors by analogy to OR83b. The assembly of IRs in multimeric ion channel complexes could create an enormous combinatorial diversity. In addition, co-operativity among sub-units could provide a mechanism for tuning channel activity to a narrow range of ligand concentrations [44]. Figure 1.5 depicts a summary of insect olfactory receptors.

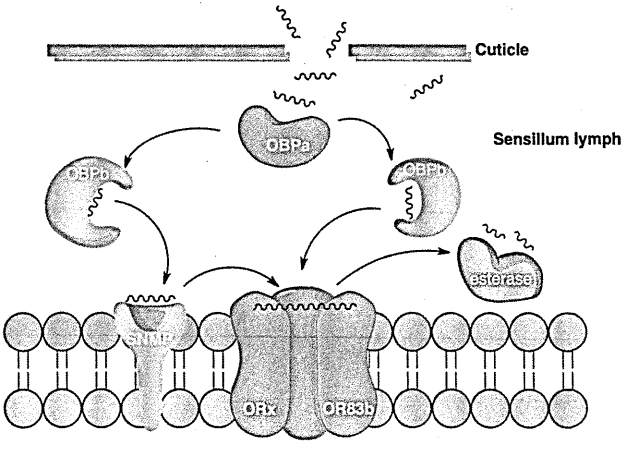



**Figure 1.5** **A schematic representation of receptor events**

OBPa: inactive form; OBPb: active form; OBPb-ligand can either react with the ORx-OR83b complex or with SNMP which then induces a conformational change of the ORx-OR83b complex; IRs are not shown

## 1.4 Odorant Binding Proteins

### 1.4.1 Diversity and Classification

Insect OBPs are highly abundant in the sensillar lymph of the antennae. Their sizes range between 15 and 20 kDa and they consist of α-helices which fold into a globular water soluble bundle. There is no sequence or structural homology between insect and vertebrate OBPs. The first insect OBP was discovered in the early '80s in the giant moth *Antheraea polyphemus* [26]. This protein was shown to bind radiolabeled pheromones and was classified as pheromone binding protein (*Apol*PBP). Since that time, the number of identified OBPs has reached the hundreds, spanning more than ~50 insect species belonging to 10 different orders. For a protein to be classified as OBP two additional criteria need to be met. First, it is important to demonstrate the ubiquitous expression of the OBP in olfactory tissues and second, its ability to bind odorants. However, ligand binding is a necessary but not sufficient condition to establish olfactory function. A protein may be involved in a similar chemosensory function in non-olfactory tissues such as in the transport of semiochemicals for example. Bioinformatic approaches have been employed to analyse genomic data for the identification of putative OBPs. These have led to the identification of additional OBP sub-families in addition to the "classic" OBP sub-families of PBPs, GOBPs and ABPs. These include, Plus-C OBPs, having eight conserved cysteines plus a proline, Minus-C OBPs, having a signature of four conserved cysteines, and atypical OBPs having the conserved motif of the classical OBP family as well as a number of other residues that are not conserved in classical OBPs [45] [46] [47]. However, for the reasons mentioned above, putative OBPs may not always correlate to their true biological function.

OBPs share a unique signature of six cysteine residues forming three disulfide bonds. These OBPs are known as "classic OBPs". It has been demonstrated experimentally that the disulfide bridges are formed between the cysteines of the helix pairs 1-3, 2-5 and 4-6 [48], [49], [50]. The number of amino acid residues between the second and third cysteine is always three, whereas that between the fifth and sixth cysteine is always eight. This six cysteine pattern is well conserved amongst the OBPs of all insect orders. With the exception of Lepidopteran OBPs, where both PBPs and GOBPs are well conserved amongst different species (50% and 30% sequence identity, respectively) [51], OBPs are highly divergent even within the same species. For example, the amino acid sequence conservation amongst the 39 classical OBPs of *D. melanogaster* is around 10-15% and in some cases as little as 4% corresponding to the six cysteine motif. In only one case is the sequence identity unusually high (60%)[52].

Genomic analyses have shown that OBP genes are located in gene clusters suggesting that gene duplication is an important mechanism to increase diversity. The sequence divergence observed amongst the members of the OBP gene clusters indicates that they have been subject to positive selection, the result of which has been the rapid evolution and the concomitant functional diversification of these genes [53]. Although exceptions to this may be possible, phylogenetic studies support overall an evolutionary birth-and-death model whereby new members of the OBP gene family arise by tandem gene duplication and progressively diverge in sequence and function. Eventually, they can be lost either by deletion or lose their function by conversion into pseudogenes. Furthermore, the OBP gene family is subject to adaptive changes brought about by environmental shifts. Such adaptations are likely to involve the evolution of new reproductive, ecological and behavioural traits. As a consequence, the size of the

OBP gene pool is in a state of flux due to random gene gain and loss events. Such events are a major source of genetic variation and provide ample opportunities for future specific adaptations [54].

### 1.4.2 Structural features

To date, only a few structures of insect OBPs have been solved by means of X-ray diffraction and NMR. These are listed in Table 1.2. Despite the relative sparsity of available models, these structures combined with other experimental data have provided important clues to the mechanism of odorant binding and release.

The structural models listed in Table 1.2 belong to the so-called classical OBPs and were the first ones to have been solved by X-ray crystallography. They share a high degree of structural similarity despite the high level of sequence divergence amongst them. More recently, two additional *A. gambiae* OBP structures were solved crystallographically, namely *Agam*OBP7 (3r1p) [55] and *Agam*OBP47 (3pm2) [56]. These two structures represent completely different folds. The proteins belong to the C-plus class of OBPs having four and six disulfide bridges, respectively. The main structural features of the classical OBPs are presented below as they are the main focus of this work.

| Insect species | OBP/PBP | UniProt ID | PDB ID | Ligand |
|---|---|---|---|---|
| | *Agam*OBP1 | Q8I8T0 | 2erb | PEG |
| | | Q8I8S8 | 3n7h | DEET/PEG |
| *A.gambiae* | *Agam*OBP4 | Q8T6R7 | 3q8i | indole |
| | *AgamOBP7* | Q7PXT9 | 3r1v | 4-{(E)-[4-(Propan-2-Yl) phenyl]diazenyl}phenol |
| | *Agam*OBP22 | Q7PGA3 | 3l4l<br>3qme | benzaldehyde<br>cyclohexanone |
| *A.aegypti* | *Aaeg*OBP1 | Q6Y2R8 | 3k1e | PEG |

**Table 1.2 OBPs solved by X-ray crystallography (contd.)**

| Insect species | OBP/PBP | UniProt ID | PDB ID | Ligand |
|---|---|---|---|---|
| *C.quinquefasciatus* | *Cqui*OBP1 | Q8T6I2 | 3ogn | (5R,6S)-6-acetoxy-5-hexadecanolide |
| *D.melanogaster* | LUSH | OB76A | 2gte | 11-*cis*-vaccenyl acetate |
| | | | 1oof | ethanol |
| | | | 1oog | propanol |
| | | | 1ooh | butanol |
| *A.mellifera* | *Amel*ASP1 | Q9U9J6 | 3bfh | hexadecanoic acid |
| | | | 3bfb | 9-keto-2E)-decanoic acid |
| | | | 3cyz | 9-keto-2E)-decenoic acid |
| | | | 3cz1 | n-butyl benzenesulfonamide |
| | | | 3bfa | queen mandibular pheromone |
| | *Amel*OBP5 | Q8WRW2 | 3r72:A | n-butyl benzenesulfonamide |
| | *Amel*OBP14 | (Q1W640) | 3s0d | 3,7-dimethylocta-2,6-dienenitrile |
| *B.mori* | *Bmor*PBP1 | P34174 | 1dqe | bombykol |
| | | | 2p70 | bell pepper odorant |
| | | | 2p71 | iodohexadecane |
| | *Bmor*GOBP2 | P34170 | 2wc6 | bombykol |
| | | | 2wch | bombykal |
| | | | 2wcm | (10E)-hexadecen-12-yn-1-ol |
| | | | 2wcl | (8E,10Z)-hexadecadien-1-ol |
| | | | 2wcj | (10E-12Z)-tetradecadien-1-ol |
| *L.maderae* | *Lmad*PBP | Q8MTC1 | 1ow4 | 1-anilinonaphthalene-8-sulfonic acid |
| | | | 1p28 | 3-hydorxy-butan-2-one blend) |

**Table 1.2 (contd.) OBPs solved by X-ray crystallography**

**Note 1.** Entries of crystallographic and/or NMR models of apoproteins and mutants of the above species are not included
**Note 2.** OBP apoprotein models are not included

### 1.4.2.1 The 3-D structure of the classical OBPs

The classical OBPs share a domain consisting of six or more α-helices packed compactly due to the three disulfide bridges interlocking six helical chains. The first two, in sequence, disulfides have conserved domain positions whereas the position of the third is more variable. These three conserved disulphide bridges impose structural constraints in the binding cavity, the binding of ligands being made possible by side-chain movements of the amino acids that enclose the cavity. Despite their similar fold, the structures studied so far show significant variations in terms of the length, position and tilt of their α-helices, in the lengths of

their loops and the length and conformation of their C-termini. Based on their size, the structural models of the classical OBPs have been grouped into long-, medium- and short-chain [57].

The long-chain group of proteins is represented by the *Bombyx mori* pheromone binding protein (*Bmor*PBP1) which was the first structure solved by X-ray diffraction in complex with its pheromone bombykol [58]. This protein, which consists of 137 residues, has a fold that is characterised by six α-helices surrounding a hydrophobic cavity in which bombykol fits. The C-terminus of the protein consists of 12 residues in extended conformation at pH 7, which switches into a seventh helix at pH in the range of 4 to 5 [59]. It has been postulated that this conformational change is of functional significance and plays a crucial role in the release of the odorant from the binding site. The pheromone is released as the seventh helix formed by the C-terminal dodecapeptide inserts itself into the binding cavity thus expelling the ligand from it [23]. This conformational transition from the active binding mode to the inactive one is exemplified in Figure 1.6.

Under physiological conditions the switch of the disordered C-terminus stretch to an ordered, amphipathic helix may occur as the PBP approaches the negatively charged dendritic membrane at the pore cuticle [60]. A similar pattern of pheromone release by an acid-induced formation of a seventh helix involving residues at the C-terminus of the protein has been reported in the case of the PBPs of *A. polyphemus* [51], [61], [62] and *A. transitella* [63], the structures of which were solved in solution by NMR. These findings led to the hypothesis that pH-induced conformational changes at the C-terminus of the "long-chain" PBPs may be a general mechanism of pheromone release from the binding cavity.

More recently, the structure of *B. mori* GOBP2 (*Bmor*GOBP2) was solved by X-ray crystallography in complex with the pheromones bombykol and bombykal, as well as with a number of analogues [64]. The most significant difference between the structures of *Bmor*PBP1 and *Bmor*GOBP2 is observed in the C-terminus. The C-terminus is longer in *Bmor*GOBP2 and forms a regular amphipathic α-helix that is overlaid across the N-terminal helix rather than inserting itself between the first and second helices as is the case in the *Bmor*PBP1-bombykol complex (Figure 1.7). There is no evidence of the C-terminus of the *Bmor*GOBP2 inserting itself into the binding site at low pH.

To what extent the insertion of the C-terminus into the binding pocket at low pH constitutes a general mechanism for ligand release of the long-chain OBPs needs to be further appraised in the light of other parameters in addition to changes in pH. Biochemical studies on the effect of salt concentration and pH on the binding properties of *Lymantria dispar* PBPs showed that although lowering of the pH

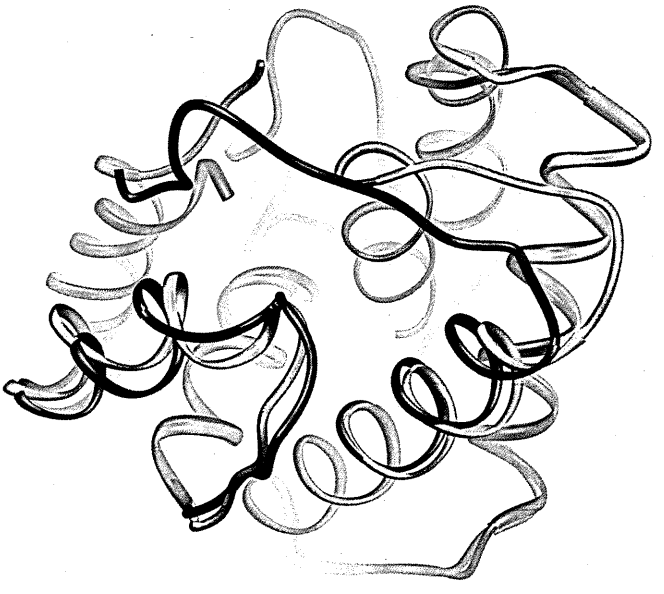


**Figure 1.6** **Superposed 3-D models of *Bmor*PBP1 at pH 7 (blue) and pH 4.5 (red)**

The unstructured C-terminus end overlying the binding cavity at pH 7 (blue) undergoes a conformational change at pH 4.5, forming a helical structure that intrudes into the binding cavity (red)

leads to dissociation of ligands from the PBPs, this effect was counterbalanced by increased concentration of KCl suggesting that the positive ion gradients in the vicinity of the sensory neuron membrane may have no effect on the release of pheromone. It was suggested that, instead, the interaction of the PBP-pheromone complex with a component of the dendritic membrane may be responsible for triggering the olfactory response [65].

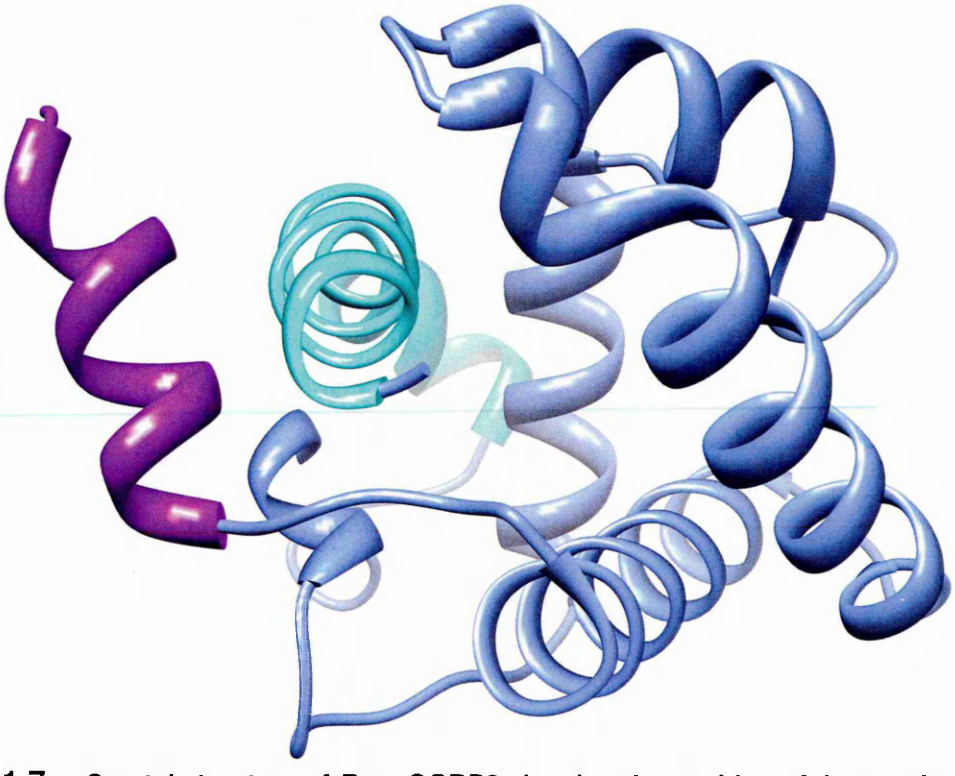


**Figure 1.7** **Crystal structure of *Bmor*GOBP2 showing the position of the terminal helices**
N-terminus: cyan; C-terminus: purple

The medium-chain sub-class of OBPs/PBPs consists of proteins of intermediate length (~125 amino acid residues) and includes as members the mosquito OBPs of *A. gambiae* (*Agam*OBP1), *A. aegypti* OBP1 (*Aaeg*OBP1), and *C. quinquefasciatus* (*Cqui*OBP1), the *A. mellifera* ASP (*Amel*ASP1) and the *D. melanogaster* OBP76a ( LUSH). The carboxy terminal ends of these proteins are also in extended conformation but shorter than that of *Bombyx mori* PBP, and are not long enough to undergo a pH dependent conformational change leading to

helix formation and displacement of the bound ligand. The OBPs of the three mosquito species share over 80% amino acid sequence identity and have strikingly similar structures. All three have been crystallised as dimers suggesting that dimerisation may play a physiologically relevant role, although this could well be a solid-state artifact due to the high OBP concentration used in crystallisation. *Agam*OBP1 has been crystallised in complex with the repellent *N,N*-diethyl-*m*-toluamide (DEET) [66] and with a 195 polyethylene glycol (PEG) molecule [67]. *Aaeg*OBP1 was crystallised also with a PEG molecule occupying a long tunnel formed by the crystallographic dimer much the same way as in the AgamOBP1/ PEG model [68]. *Cqui*OBP1 was co-crystallised with the oviposition pheromone (*5R,6S*)-6-acetoxy-5-hexadecanolide (MOP)[69].

The crystal structures of these OBPs are characterised by monomeric sub-units with a central cavity interconnected by means of a hydrophobic tunnel running through the dimer interface. The crystal models of *Agam*OBP1 and *Aaeg*OBP1 with PEG show the latter running through the interconnected cavities and exiting through openings formed by helices 1, 3 and 4 of the two sub-units. *Cqui*OBP1, likewise, has been crystallised in dimeric form with two bound molecules of MOP, one in each monomeric sub-unit (Figure 1.8). Similarly to *Agam*OPB1 and *Aaeg*OBP1, a hydrophobic tunnel runs through the dimer interface. MOP makes numerous hydrophobic and van der Waals interactions inside the binding cavity with the lactone/acetyl ester "head" of the ligand occupying approximately half of the central cavity of the receptor, whereas the long lipid "tail" of the pheromone is buried in the tunnel formed by helices 4 and 5. The crystal structure of 3ogn also reveals that the electron density around the lactone ring is not very well defined suggesting that this part of the ligand can attain different conformations inside the cavity. The binding site thus appears not to impose strict conformational

constraints on the ligand. This relative plasticity of the OBP binding site is not a distinctive feature of *Cqui*OBP1 alone and has been observed in the case of *Bmor*PBP1 as well [70].

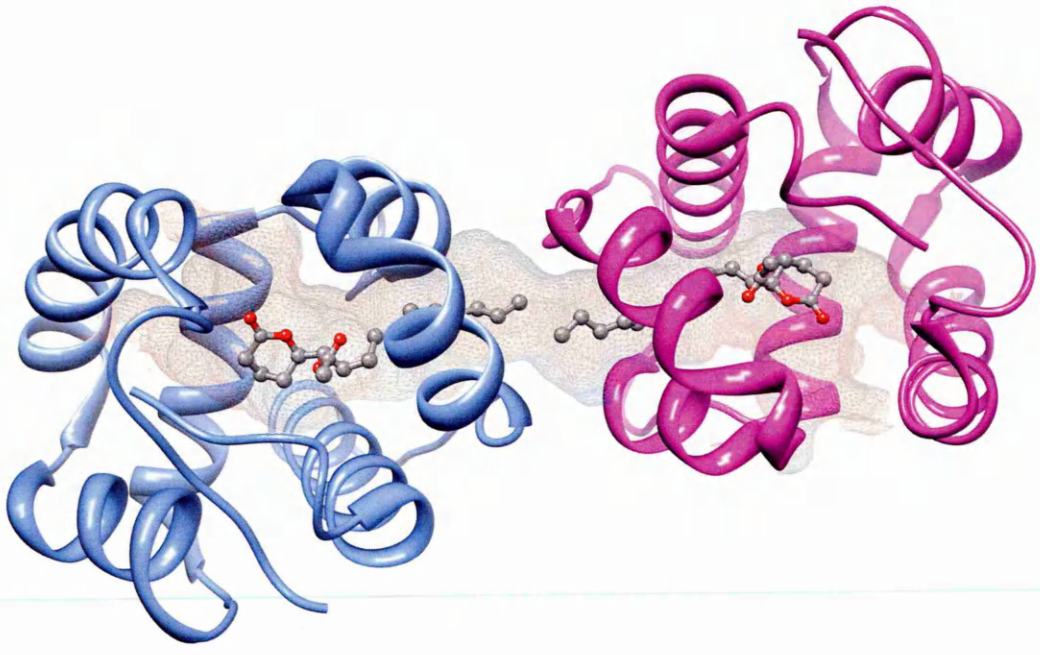


**Figure 1.8** **Crystal structure of *Cqui*OBP1 with the oviposition pheromone MOP**
The tunnel running through the dimeric interface is depicted as mesh

The *Agam*OBP1-DEET structure represents the first example of an odorant repellent in complex with an OBP. Two ligand molecules, one per monomeric subunit, are shown to bind at the end of the long tunnel interconnecting the two subunits making numerous non-polar interactions and forming a hydrogen bond to a conserved water molecule which, in turn, interacts with conserved residues in its vicinity (Trp114 and Cys95 or Gly92). This hydrogen-bond network is perceived to be critical for the recognition of the repellent [66].

Another structural feature of the mosquito OBP models merits special attention. The C-terminus of these OBPs is locked onto helices 1 and 3 through hydrogen bond interactions involving the terminal carboxylate forming a “lid” on the central

cavity. In the case of *Cqui*OBP1, this lid makes extensive contacts with both the lactone and acetyl ester groups of MOP and may thus play a functional role in constraining the ligand within the protein cavity. It has been suggested that the "lid" is pH sensitive and a drop in pH may result in the disruption of the hydrogen bond network that holds it in place resulting in the release of the ligand from the complex.

In *Dmel*OPB76a (LUSH), the C-terminal end of medium-chain OBPs has been shown to play a different role and be critical in triggering the firing of pheromone-sensitive neurons. The crystal structure of the LUSH has been solved in complex with short-chain n-alcohols [71], in apoprotein form [72] and, more recently, in complex with the pheromone 11-*cis* vaccenyl acetate (cVA)[18]. In LUSH the C-terminal end folds into the core and forms part of the binding pocket which is the reverse configuration of the one observed in *Bmor*PBP1, in which the C-terminal tail packs outside the binding pocket at neutral pH. A comparison of the LUSH-cVA structure with the apoprotein and the alcohol-bound forms showed that unique conformational changes occur upon cVA binding. These changes, which are absent in the apoprotein and LUSH-alcohol complexes, are brought about by interaction with Phe121 located at the C-terminal end of LUSH. cVA is bound in two distinct conformations; in both of them the acetate moiety of the pheromone forms polar interactions with Ser52 and Thr57, whereas the remainder of the alkyl chain of the molecule adopts anti-parallel conformations in the two monomers of the crystallographic asymmetric unit. However, both conformers interact with Phe121 and this interaction mediates conformational shifts in other amino acids in the C-terminal region (Gln120, Asp118), which in turn bring about the disruption of a salt bridge between Asp118 and Lys87 present in both the alcohol-bound and apoprotein structures. In addition to this change there appear to be other

conformational changes in other parts of the protein and, in particular in the loop between residues 116-118, indicating that cVA binds to LUSH in multiple modes that do not result in one unique protein conformation. However, as Phe121 seems to be involved in all conformational changes the authors proceeded to mutate this residue in an attempt to evaluate the effect of different size amino acid side chains at this position and whether protein conformational changes are physiologically meaningful. One particular mutation ($LUSH^{D118A}$) results in a conformation in which the Asp118-Lys87 salt bridge is absent. It was shown that $LUSH^{D118A}$ was able to trigger robust action potentials in T1 neurons in the absence of the pheromone. This combined with the fact that cVA in the absence of LUSH results in action potential of much lower intensity was taken as evidence that it is the protein-ligand complex that is responsible for the transmission of the pheromonal signal.

The structure of *Amel*ASP1 has been determined in the apo form as well as in complex with the main component of the queen mandibular pheromonal mixture, 9-keto-2(E)-decanoic acid (9-ODA) and with non-pheromonal compounds [57]. It was shown that the C-terminal tail of the protein is locked through a hydrogen bonding network involving the last main-chain residue, the ligand and Asp35. The binding pocket collapses in the absence of the ligand at pH 5.5. or even in the presence of the ligand at pH 7.0 It has been suggested that the ionised state of the Asp35 side-chain has a repulsive effect on the C-terminus causing release of the ligand at pH 7.0. This repulsion is counterbalanced in the presence of ligand at pH 5.5 [57]. It seems therefore that members of this structural sub-class of OBPs bind ligands in a pH dependent manner but their behaviour does not follow a single rule. It is noted that in the case of *Amel*ASP1-9-ODA complex, dissociation of the pheromone occurs at physiological pH contrary to the situation observed in the

case of the other members of this structural sub-class of proteins in which it occurs at acidic pH.

The short-chain group is represented by a single member, namely the PBP of the cockroach *L. maderae*. The protein has been crystallised in the apo-form as well as in complex with a component of the pheromonal blend, 3-hydroxy-2-butanone (3H2B) and with the much larger fluorescent reporter amino-naphthalene sulfonate. The crystal structure of this protein consisting of 118 residues was the first to be solved [73]. Unlike the medium- and long-chain OBPs this protein lacks the hydrophobic C-terminal peptide and presents a relatively hydrophilic binding pocket open to solvent delineated by six helices (Figure 1.9). Although the binding cavity walls and mouth is lined mainly with hydrophobic residues, it contains a significant number of polar and charged residues and is thus able to bind small hydrophilic ligands such as 3H2B.

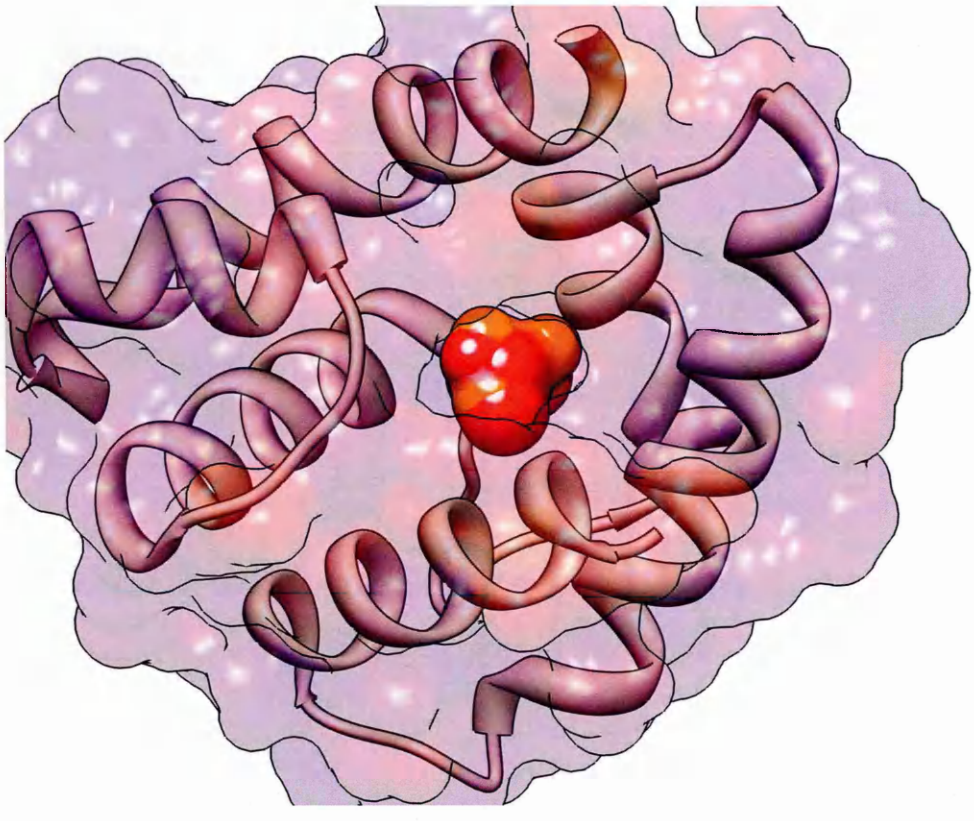


**Figure 1.9** ***L. maderae*** **PBP in complex with 3H2B**

### 1.4.2.2 Dimerisation of OBPs

It has been shown that insect OBPs can form dimers under physiological conditions [74]. The oligomerisation state of OBPs has implications regarding their function. The literature on the matter is not conclusive. The PBP of *Bombyx mori* (*Bmor*PBP1) has been reported to exist both as a monomer [75] and as a dimer [76]. An early biochemical study on the oligomeric state of *Bmor*PBP1 confirmed a dynamic monomer-dimer equilibrium in which the dimeric form is favoured at slightly acid, neutral and basic pH. Below pH 5.5 the dimer undergoes a conformational change and dissociates to a monomeric form [77]. Furthermore, in the case of *Lymantria dispar*, pheromones PBP1 and PBP2 form dimers and higher order oligomers in native gels which can bind pheromones at least three times more strongly relative to the monomeric forms [21]. A possible explanation of these observations is that OBPs may serve a dual function. In their monomeric form they may act as general 'transporters' of semiochemicals and, in their oligomeric form, as 'scavengers'. Evidence of such functional duality comes from a recent kinetic model of *Lymantria dispar* PBP2 monomer-multimer equilibria. This study provides additional evidence for the existence of PBP multimers in solution and that multimerisation state increases over a long time (hours) in the presence of ligand [78].

The formation of homodimers has also been shown in crystallographic models in which the biological unit is shown to contain two monomeric chains. Analysis of several OBP crystallographic models shows that the dimer interface in some structures is not well defined, as for example, in the crystal structure of the *Bmor*PBP1-bombykol complex, where the dimer interface is far too small and non-symmetrical [58]. In other cases, such as in the crystal structure of *Agam*OBP1, the dimeric interface is clearly defined but not structurally conserved [67].

Dimerisation could be favoured in cases where the concentration of the OPB in the perireceptor space is high. Although, in most cases, the actual physiological concentration of OBPs is not known, there is evidence that in some cases it can be as high as ~10 mM [79]. In principle, oligomerisation of OBPs could provide a mechanism whereby these polypeptides bind more that one ligand in different parts of the molecule. Thus an OBP could bind ligands in a well defined binding pocket as well as the oligomer interface. The ligands could bind simultaneously or at different times during the flight of the insect and the encounter of different odorants [67].

Heterodimerisation has also been shown to occur in i*n-vitro* experiments involving *Agam*OPB48 being co-expressed in the same cells as the *A. gambiae* OS-E and OS-F genes. This protein, which belongs to the Plus-C group of OBPs, is primarily expressed in female antennae and is down-regulated after a blood meal. Co-immunoprecipitation and chemical cross-linking studies showed that *Agam*OBP48 is capable of forming homodimers, heterodimers and higher order complexes with the classical OBPs OS-E and OS-F [80]. Heterodimerisation could provide a mechanism whereby OBP paralogues arising from a gene duplication event could be fixed in the genome as this could, theoretically, increase the repertoire of ligand binding sites beyond those of the individual monomeric forms.

### 1.4.3 OBP-ligand binding

Monitoring of OBPs/PBPs binding to specific semio-chemicals and pheromones has been an essential tool in deciphering whether these proteins can selectively bind and differentially recognise compounds of the olfactory repertoire. Ligand binding affinities, as measured by the dissociation constant of the protein-ligand complex, are weak (μM-nM range). Numerous binding assays have been

conducted in several insect species, most of which involved proteins expressed in heterologous systems. Recombinant OBPs/PBPs are deemed to have the same functional and structural properties, much like the native proteins as

- they are not subject to post-translational modifications;
- they are expressed in the cellular periplasm in conditions allowing the formation of the signature disulphide bonds;
- their CD spectra are consistent with the secondary structure predictions;
- their crystal structures have been generated by the same heterologous expression procedures.

Different approaches have been followed to measure binding and dissociation constants in OBP/semiochemical complexes. Originally, antennal extracts were analysed with tritium labelled pheromone [26] and also with radioactive photoaffinity labels [81]. These studies, on occasions, produced erroneous predictions regarding the binding sites of the tested ligands. For example, binding of photoactive compounds on BmorPBP1 occurred on the protein surface rather than in the internal binding site [79].

More recent studies have made use of fluorescent binding assays [82] [83], immunochemistry techniques [84], two-phase-binding assays [64], as well as hybrid techniques, such as “cold”-binding [85] and calorimetry volatile odorant-binding assays [86]. In addition, X-ray and NMR models and electrospray mass spectrometry (ESI-MS) have provided useful insights on the binding mode of several odorant molecules and pheromones.

The results obtained from these studies indicate broad specificity binding not only for OBPs but also for PBPs. For example, fluorescent displacement binding assays, involving the PBPs of *Mamestra brassicae* (*Mbra*PBP1), a hexa-mutant of *MBra*PBP1, *Antheraea polyphemus* (*Apol*PBP1), and *Bombyx mori* (*Bmor*PBP1), their natural pheromones, the fluorescence probe 1-aminoanthracene (AMA) and several fatty acids, provided no evidence of discrimination of the physiologically relevant pheromones [82]. Similar studies, involving affinity displacement of a fluorophore from the binding site of the PBP of *Leucophaea maderae* (*Lmad*PBP) by a pheromonal blend showed that some components of the blend were able to displace the fluorophore whereas others not [87]. Furthermore, the components of the pheromonal blend (3-hydroxy-butan-2-one and butane-2,3-diol) that were able to displace the fluorophore are water soluble molecules in contrast with most natural pheromones, which are almost exclusively very lipophilic compounds. The findings of this study do not exclude the possibility that PBPs may function as filters being able to discriminate some ligands from others, or alternatively, some PBPs are more selective than others. The PBP of the silk moth *B. mori* (*Bmor*PBP1) was shown to bind, in addition to its natural pheromone, bombykol, non-pheromone compounds such as (10,12)-hexadecadiyn-1-ol [88]. This study, which employed high-throughput ESI-MS analysis for the *Bmor*PBP1/ligand complexes, revealed that whereas the protein tolerated different ligands, binding affinity increased with the degree of unsaturation, and the proteins showed selectivity for ligands with polar moieties at the terminus. It was shown to discriminate strongly against bombykal, which is also a pheromone component, as well as against 1-iodohexadecane, a compound that was found from crystallographic studies to form a complex with *Bmor*PBP1.

Experiments with the pea aphid *A. pisum* OBP (*Apis*OBP3) demonstrated that the latter exhibits a broad spectrum of binding but, nevertheless, detects and recognises the alarm pheromone (*E*)-β-farnesene [89]. The PBP of *A. mellifera* was crystallised from the four-component queen mandibular pheromone mixture in complex with the main pheromonal component only (*E*)-9-keto-2-decenoic acid [57].

It has been suggested that under competitive conditions OBPs may discriminate better the individual components of pheromonal blends. Competitive binding assessment using pheromonal blends have shown that in the case of *A. polyphemus*, *Apol*PBP1, incubated with equal amounts of each of the three pheromonal constituents of the blend, binds the main pheromone component preferentially to the other two [90]. Similarly, the PBP of the gypsy moth, *Lymantria dispar*, binds its natural pheromone (*7R, 8S*)-2-methyl-7,8-epoxyoctadecane, with some discrimination which increases in the presence of its known antagonist, (*7Z*)-2-methyloctadec-7-ene. Binding enhancement also occurs at high ligand:PBP ratios, suggesting that the PBP may function both as a pheromone transporter and as scavenger [16]. Competition experiments, involving three PBP sub-types each of *A. polyphemus* and *A. pernyi* in solution with tritium-labelled pheromone components (*E,Z*)-6,11-hexadecadienyl-acetate and (*E,Z*)-6,11-hexadecadienal, showed binding preference for each of the PBP sub-types for as specific pheromone component [91]. PBP from both species bind strongly decyl-thio-1,1,1-trifuoropropanone (DTFP) a known inhibitor of sensillar esterase. DTPF caused a marked reduction of electrophysiological responses to each of the two pheromone compounds. As this reduction could not be attributed to inhibition of the esterase, it was suggested that it was either due to the interaction of DTPF

directly with the ORs or to binding to the PBP thus preventing the formation of the pheromone-PBP complex and the concomitant stimulation of the receptors [92].

The conclusions of most of the studies mentioned above support the idea that OBPs/PBPs contribute to odour discrimination, acting either as selectivity filters for specific types of compounds, or by inducing receptor activation. The reason why binding affinities toward ligands are not very strong may be explained by the fact that OBP/PBP concentration in the sensillar lymph is very high (~10 mM) [26]. PBP:ligand dissociation constants in the range of 10 mM would be enough to keep 99% of the ligand bound to the PBP even if the concentration of the latter is an order of magnitude less than what has been reported [93].

## 1.5 Motivation

The disruption of olfactory behaviour of agriculturally important insect pests and/or vectors for the transmission of deadly disease is a legitimate biotechnological target. Despite the ambiguity of the biological function of OBPs, there is an adequate body of evidence suggesting that, at least some of them, play an important role in olfactory receptor activation. At the molecular level, the weak selectivity of OBPs presents a major challenge in determining the residues from which this selectivity originates. As progress in the field of insect olfaction gains pace, bioinformatics and modelling approaches are likely to play a pivotal role in the structure-based design of potential modulators of insect behaviour.

Recent examples of computational studies involving OBPs, include the modelling of *Agam*OBP1-indole binding [94]. In this work molecular mechanics calculations were used to optimise indole conformation in the binding pocket of *Agam*OPB1. Homology modelling of *Agam*OBP3 and *Agam*OBP4 has been used to identify

structural similarities of these proteins with *Agam*OPB1 [95]. Structure-based modelling and docking simulations have also been performed to facilitate the design of novel mosquito repellents [66]. The ligand specificity of numerous volatile odour compounds towards an OBP of *Locusta migratoria* (*Lmig*OBP1)was investigated by means of docking simulations on a homology model of *Lmig*OBP1 [96]. Molecular dynamics (MD) simulations were also used to decipher the mechanism and energetics of bombykol uptake and release from *Bmor*OBP1. These simulations predicted the presence of two gateways allowing the random diffusion of bombykol to, and release from, the binding cavity of *Bmor*OBP1.

Two other important examples of computational studies on insect OBPs include the work of Golebiowski and co-workers, who have used thermodynamic integration methods and free energy calculations to determine the differential selectivity of *Bmor*PBP1 towards bombykol and bombykal [97]. The authors used per residue decomposition of the interaction energy to identify residues responsible for the ligand selectivity within the binding pocket of the protein. They found that *Bmor*PBP1 is able to bind both compounds, although its affinity to bombykol was somewhat higher. The results of this study contradict an earlier and more extensive MD study, which revealed that both compounds bind to *Bmor*PBP1 with similar activity but bombykal forms different polar interactions from bombykol due to the different chemical preferences of the carbonyl group [98]. Both studies demonstrated that the binding cavity of *Bmor*PBP1 has remarkable plasticity and can accommodate bombykol in significantly different conformations. In fact, docking simulations revealed that bombykol can bind to the protein in many-fold conformations with various hydrogen bonds to the protein and with similar binding energies.

Free energy calculations are now routinely used to predict binding affinities of small drug-candidate compounds and to evaluate relative stabilities of protein structures. End-point implicit solvent approaches are often preferred for such calculations to the rigorous but computationally expensive free energy perturbation (FEP) [99] and thermodynamic integration (TI) methods [100]. Probably the most widely used end-point, implicit solvent free energy method is the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) approach [101]. MM-PBSA has been shown to be of a relatively low computational cost, provide a trade-off between efficiency and accuracy, and be more accurate in predicting binding poses and binding affinities than simple scoring functions. Another advantage of the method is that it can provided detailed insights of atomistic interactions in a macromolecular complex by means of a per-residue decomposition of free energy.

The study of insect OBPs as potential biotechnological targets for insect control is confronted with a number of challenges, including:

- the very large number of putative OBPs. Inevitably work will have to focus on determining those to be most abundantly expressed in olfactory organs;

- the dearth of X-ray or NMR structures in complex with biologically meaningful odourant molecules. Homology modelling, docking simulations and free-energy calculations can provide valuable leads on compounds that could possibly disrupt olfactory function;

- little is known about the interaction of OBPs with specific receptors and the mode of delivery of odourant molecules to the receptor. Once such

information emerges from biochemical studies, it may be possible to use computational methods to predict the means to disrupt OBP-OR interactions.

The work at hand represents the first comparative study of binding affinities of homologous moth and mosquito OBPs through the concerted use of different computational methods. The overarching objective of this work is to use molecular modelling, docking, MD simulations and MM-PBSA calculations in order to gain further insights as to whether OBPs are selective in the binding of pheromones and/or putative odorant molecules. More particularly, this work focuses on structural bioinformatics approaches to:

1. model two OBPs from an important agricultural pest, *M. sexta*, and predict the binding affinity and selectivity of these OBPs to the pheromonal components bombykol and bombykal;

2. assess the differential ligand-binding capacity of homologous OBPs from three mosquito species known to be involved in the transmission of devastating diseases;

3. identify "hot spots" for ligand binding for structure-based design of compounds with a potential to disrupt olfactory behaviour.

*Chapter 2*

# *Theoretical Background and Methods*

# 2.1 Theoretical background

## 2.1.1 Introduction

In this work the interaction of insect pheromones and other putative odorant molecules with PBPs and GOBPs is studied by molecular modelling techniques, primarily docking and MD. In recent years these techniques have become a staple in structure-based drug design.

The objective of this study is to use the three-dimensional atomic coordinates of the targeted OBP/PBP to calculate the binding energy of a putative odorant or pheromone molecule and/or rank such molecules according to their predicted affinities to the target. Much like structure-based modelling, two overarching problems need to be addressed, namely (a) the accurate prediction of binding energies, and (b) the effective sampling of the available conformational space by the interacting molecular species.

Given the fact that differences in the binding free energy of a protein-ligand complex of only 1.4 kcal/mol result in a 10-fold change in binding affinities (see Section 2.2.1, page 76), computational methods strive to be at least as accurate as experimental methods with binding affinities to within 1 kcal/mol. Considering that for a medium-sized protein ligand complex the total molecular mechanics energy can be several thousand kcal/mol, the calculation of free energies within 1 kcal/mol error margin remains a goal yet to be achieved by current computational methods. A second major challenge of modeling protein-ligand interactions is conformational sampling. This is illustrated by the fact that for a given ligand the number of accessible conformations ($N$) is related to both the number of

conformations per rotatable bond ($N_i$) and the number of rotatable bonds ($n$) through equation:

$$N = N_i^n \qquad 2.1$$

Thus, assuming only three conformations per rotatable bond, the number of accessible conformations increases from 243 for a ligand with five rotatable bonds to 3.5 x $10^9$ for a ligand of 20 such bonds. The extent of the sampling problem becomes all too apparent if the number of rotatable bonds of the protein are to be considered as well.

Computational methods based on either stochastic approaches such as Monte Carlo simulations or deterministic approaches such as MD are now regularly used to address the sampling problem and gain insights of biomolecular interactions at the atomic level. Less computationally expensive docking methods are also used routinely in computer-aided drug design. An overview of MD and docking methods is presented below as these methods are the main computational approaches taken in this work.

### 2.1.2 MD: an overview

The essence of MD is encapsulated in the statement "*numerically solve the N-body problem of classical mechanics*". By examining the motions of individual particles, MD attempts to decipher the equilibrium properties and collective dynamics of molecular motions. The computational implementation of MD can be seen as "*statistical mechanics by numbers*" [102]. It enables the simulation of the motion of a collection of particles under the influence of a particular potential energy function over time according to Newtonian physics. MD provides snapshots of the behaviour of a system over space and time, thus overcoming the limitations

inherent in viewing biomolecular systems in average frozen views such as those obtained from crystallographic models. The information derived from MD simulations includes molecular geometries and energies, mean atomic fluctuations of the system as well as local ones, and energies and configurational changes due to covalent and non-covalent binding.

The theoretical foundation of molecular simulations is given in textbooks [102], [103] and review articles [104], [105]. The first MD simulations of a protein using the so-called hard-sphere model were conducted during the 1950s [106], whereas more sophisticated MD simulations using physics-based first-principles followed in the late 1970s [107]. MD simulations are now commonly used in the fields of protein chemistry and biophysics and apply to processes that have time scales ranging from femtoseconds to hours and cover an extensive range of amplitudes and energies (Table 2.1). MD is used to study both time-dependent (kinetic) and/or time-independent molecular properties that approach experimentally measurable thermodynamic averages. Recent advances in the state of the art of atomic-level simulations have enabled the study of hitherto intractable biological problems on timescales as long as milliseconds, such as protein folding, ligand binding and membrane transport [108].

| Type of motion | Example | Functionality examples | Time / amplitude scales |
|---|---|---|---|
| Local | Atomic fluctuation<br>Side chain motion | Ligand docking flexibility<br>Diffusion pathways | $10^{-15}$ - $10^{-12}$ s<br>< 1 Å |
| Medium Scale | Loop motion<br>Helical motion | Active site conformations<br>Binding specificity | $10^{-9}$ - $10^{-6}$ s<br>1 - 5 Å |
| Large Scale | Subunit motion<br>Domain motion | Allosteric transitions | $10^{-6}$ - $10^{-3}$ s<br>5 - 10 Å |
| Global | 2ry structure transitions<br>Folding / unfolding<br>Subunit association / dissociation | Protein functionality | $10^{-3}$ - $10^{4}$ s<br>> 5 Å |

**Table 2.1** **Biomolecular motions: ranges of characteristic time scales**

A brief summary of the physical principles underpinning MD simulations is presented below.

### 2.1.3 Statistical Mechanics: Phase Space and Ensembles

MD simulations study the classical motion of a many-body system and as such generate information at the microscopic (atomistic) level. The energetic relationships between different conformational states of a biomolecule are determined by thermodynamic principles, whereas transformations from one state to another are described by kinetics.

In classical mechanics, the state of a system consisting of $N$ particles, at any given time, is described by the position and velocity vectors of each constituent particle. The Hamiltonian formalism of classical mechanics uses instead generalised coordinates and momenta to describe the mechanical state of a many-particle system. At any time instant, the set of momenta $p^N$ and positions $r^N$ of each of the $N$ particles defines a unique point in a $6N$-dimensional space, known as the *phase space*, each point of which corresponds to a microscopic state of the system. The evolution of the system through phase space is determined by Hamiltonian dynamics, that is the system obeys equations:

$$\frac{dr_i}{dt} = \frac{\partial H}{\partial p_i} \text{ and } \frac{dp_i}{dt} = \frac{\partial H}{\partial r_i} \qquad 2.2$$

where $H$ is the Hamiltonian operator and $i$ varies from 1 to $N$. According to Newton's second law of motion, if the positions and velocities of all particles in the system are known at any point in time then it is possible to predict the past evolution of the system from its initial state leading to that point in time, as well as its future evolution from that point onwards. In other words, classical mechanics

encodes all the information that is needed to predict the state of a system at any instant in time.

However, any effort to apply the microscopic laws of motion to predict the properties of a macromolecular system is confronted with two problems. An analytical solution is not possible, firstly because the forces characterising real-world systems are highly nonlinear in nature, and secondly because the number of degrees of freedom ($6N$) is enormous considering that 1 mole consists of 6.022 x $10^{23}$ particles. Neither is a numerical solution of the equations of motion feasible as storing a single phase space point for a system of $10^{23}$ particles is beyond computational capabilities available presently or at any time in the foreseeable future.

Statistical mechanics addresses this conundrum by providing a reconciliation of macroscopic thermodynamics with the microscopic laws of motion. This is done by evoking the concept of the *ensemble*. This concept, which is a cornerstone in the foundation of statistical mechanics, is used to describe a collection of systems characterised by the same microscopic interactions and constrained by the same set of macroscopic properties (e.g. same volume, number of particles and total energy). If each system is allowed to evolve from a different starting state according to the microscopic laws of motion, at any given time instant, every system will be at a different microscopic state. The macroscopic properties of the ensemble can therefore be derived by performing averages over its constituent systems.

Ensembles are conceptual devices and can be defined for a variety of thermodynamic states examples of which are shown in Table 2.2. These ensembles, known as *equilibrium ensembles*, are appropriate to describe time-

independent properties of a system. It is therefore possible to freeze the ensemble at any point in time and perform the average over the collection of systems comprising it.

| Ensemble | Thermodynamic constraints | Equilibrium states |
|---|---|---|
| microcanonical (*NVE*) | constant $N, V, E$ | maximum entropy ($S$) |
| canonical (*NVT*) | constant $N, V, T$ | minimum Helmholtz free energy ($A$) |
| isothermal-isobaric (*NTP*) | constant $N, T, P$ | minimum Gibbs free energy ($G$) |
| grand canonical | constant $\mu, V, T$ | maximum pressure x volume ($PV$) |

**Table 2.2** **Principal ensembles of statistical thermodynamics**

The ensemble average is given by:

$$\langle A \rangle_{ensemble} = \iint dp^N \, dr^N A\left(p^N, r^N\right) \rho\left(p^N, r^N\right) \qquad 2.3$$

where $A$ denotes a macroscopic equilibrium observable, and $A(p^N, r^N)$ a microscopic phase space function used to calculate $A$. Thus the phase space function must be determined for any single member of the ensemble at the time instant that the ensemble was frozen. The double integral sign is written for convenience as there should be $6N$ integral signs on the integral corresponding to the $3N$ Cartesian coordinates and $3N$ components of the momenta of the $N$ particles comprising the system; $A(p^N,r^N)$ is the property of interest expressed as a function of momenta $p$ and positions $r$; $\varrho(p^N,r^N)$ is the equilibrium *probability density,* of the ensemble, that is the probability of finding a configuration with positions $r^N$ and momenta $p^N$. It is given by:

$$\rho\left(p^N, r^N\right) = \frac{1}{Q_{NVT}} \exp\left(-H(p^N, r^N) \; / k_B T\right) \qquad 2.4$$

where $H(p^N,r^N)$ is the Hamiltonian of the system, $Q_{NTV}$ is the partition function, $k_B$ is

Boltzmann's constant and $T$ is the temperature. While it is not possible to solve the equations of motion for a $10^{23}$ particle system, it is possible to calculate them numerically for systems with number of particles in the range of $10^2$ - $10^9$ depending on the complexity of the system and subject to the system being *ergodic* (see below). Consider, for example as a case in point, a microcanonical ensemble with energy $E$. This ensemble, which does not exchange energy with its surroundings, consists of all microscopic states on the constant energy hypersurface $H(p^N,r^N) = E$. The equations of motion conserve the total energy, and therefore $H(p^N,r^N)$ is constant. Hence a trajectory computed by means of the Equations 2.2 will generate microscopic configurations belonging to the microcanonical ensemble. If the trajectory is allowed to evolve for an infinite amount of time, it will sample all the configurations on the constant energy hypersurface; such a system is said to be *ergodic*. The behaviour of systems having this property is summed up by a fundamental postulate of statistical mechanics, known as the *ergodic postulate*. According to this, if a system is allowed to evolve indefinitely in phase space, it will eventually visit all the states that have a non-zero probability of existence and the time-average of any property $<A>$ of the system is equal to the ensemble average of that property.

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} \qquad 2.5$$

The time-average is given by:

$$\langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\left(p^N(t), r^N(t)\right) dt \qquad 2.6$$

where $\tau_m$ is a microscopic time scale, $\tau$ is orders of magnitude larger than $\tau_m$ thus smoothing out fluctuations on the microscopic time scale, $t$ is the simulation time and $A(p^N,r^N)$ is the instantaneous value of $A$.

In MD simulations, the choice of an appropriate energy function to reproduce faithfully the thermodynamic properties of interest needs to be complemented by a method to generate the ensemble of configurations that will be used to calculate the configurational integral. In an MD calculation, given a set of starting conditions, it is possible to solve Equation 2.6 numerically using a particular numerical integrator. The integrator generates phase space vectors at discrete times that are multiples of a time discretisation parameter or time step. The integrator is applied iteratively and the ensemble average of a property is then related to the discretised time average by:

$$\langle A \rangle_{ensemble} = \langle A \rangle_{time} = \lim_{\tau \to \infty} \frac{1}{\tau} \int_{t=0}^{\tau} A\left(p^N(t), q^N(t)\right) dt \approx \frac{1}{M} \sum_{n=1}^{M} A\left(q^N, p^N\right) \qquad 2.7$$

where $M$ is the number of time steps in the simulation.

### 2.1.4 Potential Energy Functions

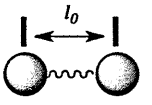
A biomolecular system is defined by the relative coordinates of its constituent atoms. Intermolecular and intramolecular interactions are described by an appropriate energy function, or force field. The quality of a molecular model is critically dependent on the choice of an appropriate force field. Force fields (also known as molecular mechanics) methods calculate the energy of a system solely as a function of the atomic coordinates, ignoring electronic motions. Numerous force fields have been developed to represent a variety of organic compounds and biomolecules, such as AMBER [109], MM4 [110], NAMD [111], GROMOS [112], CHARMM [113], and OPLS [114]. These force fields rely on simple functional forms and sets of parameters derived from experimental and/or quantum mechanical studies. A typical functional form is given by the following equation:

$$E_{MM} = E_{bond} + E_{angle} + E_{dihedral} + E_{impr-dhd} + E_{non-bonded} \quad 2.8$$

Equation 2.8 shows that the force field is additive, that is each term in the right hand of the equation is independent of the others. Although this is usually considered to be a reasonable approximation of the potential-energy landscape, it is not flawless [104].

Bonding contributions to $E_{total}$ account for interactions between two atoms (bond stretching) are described by harmonic potentials as shown in equation 2.9:

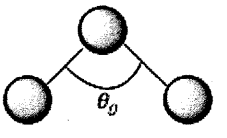
$$E_{bond} = \sum_{i=1}^{N_{bonds}} k_{l,i}\left(l_i - l_{0,i}\right)^2 \quad 2.9$$




where $k_{l,i}$ is a constant for a given bond type, $l_i$ is the bond length and $l_{0,i}$ is the reference bond length.

The angle contributions to $E_{total}$ account for interactions between three atoms (angle bending) are also described by harmonic potentials:

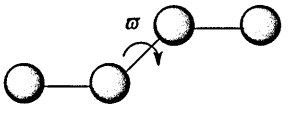
$$E_{angle} = \sum_{i=1}^{N_{angles}} k_{\theta,i}\left(\theta_i - \theta_{0,i}\right)^2 \quad 2.10$$




where $k_{\theta,i}$ is a force constant, $\theta_i$ is the value of the angle formed between the two bonds at the third common atom, and $\theta_{0,i}$ is the reference angle.

The torsional term ($E_{dihedral}$) is a four-body potential term accounting for rotations about pairs of (1-4) atoms[2] and is computed according to:

$$E_{dihedral} = \sum_{i=1}^{N_{dihedrals}} \frac{V_n}{2}\left(1 + \cos\left(n\omega - \gamma\right)\right) \quad 2.11$$

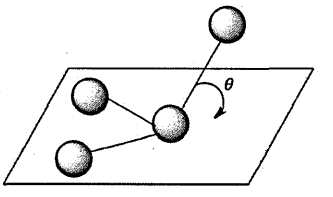



[2] interactions between the end atoms involved in a dihedral angle

where $V_n$ is an empirical term giving a qualitative indication of the relative barriers to rotation; $n$ is known as the multiplicity, its value giving the number of minimum points in the function as the bond is rotated through 360°, and $\gamma$ is the phase factor which determines the point where the torsion angle passes through its minimum value.

In order to achieve the geometry of planar groups (e.g. benzene rings) an out-of-plane bending term is used. The torsional potential ($E_{impr\text{-}dhd}$) used to maintain the improper torsion angle at 0° or 180° is:

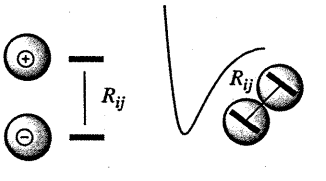
$$E_{impr-dhd} = \frac{k}{2}\theta^2 \qquad 2.12$$




where $k$ is the force constant applied to ensure planarity and $\theta$ is the angle shown in the above figure. There are other approaches to include out-of-plane bending contributions, and usually, the improper torsion term is incorporated into the 'proper' torsional term ($E_{dihedral}$).

Finally, the non-bonded potential is composed of a Coulombic and a Lennard-Jones term:

$$E_{non-bonded} = \sum_i \sum_{j>1} \left\{ \frac{q_i q_j}{4\pi\varepsilon_0 R_{ij}} + 4\varepsilon_{ij} \left[ \left( \frac{A_{ij}}{R_{ij}} \right)^{12} - \left( \frac{B_{ij}}{R_{ij}} \right)^{6} \right] \right\} \qquad 2.13$$




where the sum is over all atom pairs $i$, $j$; $q_i$, $q_j$ are the partial atomic charges, $\varepsilon_0$ is the permittivity of free space, $\varepsilon_{ij}$, $R_{ij}$ is the interatomic distance, and $A_{ij}$ and $B_{ij}$ are the Lennard-Jones well-depth energy and collision-diameter parameters respectively. The non-bonded term is also applied to 1-4 atoms but it is scaled down by an appropriate factor.

For biomolecular systems the number of pairwise non-bonded interactions can be extremely large and the inclusion of all of these in the calculation of energy would be computationally very expensive. For a system of $N$ atoms, the number of computational steps to calculate all such interactions would increase as the square of the number of atoms $N^2$. Computational cost can be reduced by calculating interactions between pairs of atoms that lie within a pre-defined cutoff distance, and ignoring interactions between pairs of atoms lying outside this range. For short-range interactions, such as van der Waals, such an approach does not introduce significant errors. However, distance cutoffs may introduce serious discontinuities and errors for long range electrostatic interactions. These problems are addressed by scaling the non-bonded terms through multiplication with a switching function [103] or, more recently, through the introduction of methods such as the particle mesh Ewald (PME) summation [115] [116].

### 2.1.5 MD simulations

MD simulations involve the study of the time evolution of a set of interacting atoms constituting a biomolecular system. The underlying methodology is the integration of the Newtonian equations of motion. Newton's second law or equation of motion relates the net force $F_i$ exerted on atom $i$ by the remainder of the system to its mass $m_i$ and acceleration $a_i$ according to:

$$\vec{F}_i = m_i\vec{a}_i = m_i\frac{d\vec{v}_i}{dt} = m_i\frac{d^2\vec{r}_i}{dt^2} \qquad 2.14$$

Therefore, velocity $v_i$ is the first derivative of the position $r_i$ with respect to time and acceleration is the second derivative of the position $r_i$ with respect to time.

The net force $F_i$ acting on atom $i$ is also related to the potential energy $U$ of the system by:

$$\vec{F}_i = m_i\vec{a}_i = -\frac{dU_i}{d\vec{r}_i} \qquad 2.15$$

Position, velocity and acceleration are therefore a function of time, and the force is related to the acceleration and the potential energy. Combining Equations 2.14 and 2.15 gives:

$$\vec{F}_i = -\frac{dU}{d\vec{r}_i} = m_i\frac{d^2\vec{r}_i}{dt^2} \qquad 2.16$$

The main objective of the MD simulation is the calculation of the net force acting on each atom over time. This information is then used to calculate the atomic positions. The total force on each atom $F_i$ is the vector sum of its interactions with all other atoms. Energies can be calculated by molecular mechanical and/or quantum mechanical methods, and consequently forces can be derived from Equation 2.16. Knowledge of the atomic forces and masses, in turn, can be used to determine the atomic positions along a series of extremely short time steps $\Delta t$ through the integration of Equation 2.16. The problem is that no analytical solutions are possible for Equation 2.16 as the atomic motions are coupled together. This problem can be overcome by employing finite difference methods for the integration of the equations of motion. All such methods assume that positions, acceleration and velocities can be approximated by a Taylor series expansion. For example, the acceleration of atoms can be determined from the force, accelerations are then combined with the positions and velocities at time t to calculate the positions and velocities at time $t+\Delta t$ by:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + d\vec{v}_i(t) + \frac{1}{2}dt^2\vec{a}_i(t)\Delta t + \frac{1}{6}dt^3\vec{b}_i(t) + \frac{1}{24}dt^4\vec{c}_i(t) + .... \qquad 2.17$$

where the first and second derivative terms are the velocity and acceleration of atom $i$, respectively (see Equation 2.14), $\vec{b}_i$ is the 3rd derivative, etc. Velocities and accelerations, as well as the higher terms $\vec{b}_i$ and $\vec{c}_i$ etc. can be approximated similarly:

$$\vec{v}_i(t+\Delta t) = \vec{v}_i(t) + dt\vec{a}_i(t) + \frac{1}{2}dt^2\vec{b}_i(t)\Delta t + \frac{1}{6}dt^3\vec{c}_i(t) + .... \quad 2.18$$

$$\vec{a}_i(t+\Delta t) = \vec{a}_i(t) + dt\vec{b}_i(t) + .... \quad 2.19$$

For any given position, the forces acting on atom ($i$) can be calculated by:

$$\vec{F}_i(t) = m_i \frac{d^2\vec{r}_i(t)}{dt^2} \quad 2.20$$

A popular algorithm for integrating the equations of motion is the leapfrog algorithm [117]. The algorithm essentially consists of three sequential steps involving:

1. calculation of accelerations at time $t$ from the force (Equation 2.14)

2. updating velocities at $t+\Delta t/2$ using:

$$\vec{v}_i(t+\frac{\Delta t}{2}) = \vec{v}_i(t-\frac{\Delta t}{2}) + \frac{d\vec{r}_i(t)}{dt} + \vec{a}_i(t)\Delta t \quad 2.21$$

3. updating positions at $t+\Delta t$ using:

$$\vec{r}_i(t+\Delta t) = \vec{r}_i(t) + \vec{v}_i(t+\frac{\Delta t}{2})\Delta t \quad 2.22$$

The leapfrog algorithm is a modification of the simple Verlet algorithm [118]. Other commonly used algorithms based on finite difference methods include the velocity Verlet method [119] and the Beeman algorithm [120]. Finite difference methods are known to produce incorrect trajectories due to the truncation of equation 2.18. A

family of integrator algorithms known as predictor-corrector methods have been developed to address this problem [121].

The order of truncated differential equations is chosen such that it balances computational speed with accuracy. However, the computational demand for solving the equations of motion is much smaller than that required for the calculation of the forces acting on all atoms within the system. Reduction of the computational cost can be achieved by using longer time steps ($\Delta t$) as this would require fewer iterations to cover a given trajectory but too long a time step can introduce discretisation errors [122].

For all algorithms, initial coordinates at time zero ($t_o$) are taken from experimental (X-ray, NMR) or homology models and velocities are initialised randomly provided that (a) the translational momentum be conserved, (b) the kinetic energy be related to the equipartition theorem (Equation 2.23), and (c) the velocities be distributed according to a Maxwell-Boltzmann or Gaussian distribution:

$$\frac{3}{2}Nk_BT = \sum_{i=1}^{N}\frac{m_i\vec{v}_i^2}{2} \qquad 2.23$$

where $k_B$ is the Boltzmann constant, $N$ the number of atoms in the system, and $T$ a user specified temperature. Initial accelerations are calculated from the initial forces or from the derivative of the potential energy (see Equations 2.14 or 2.15).

### 2.1.6 Simulation of experimental conditions

Biological reactions take place in an essentially aqueous environment with water molecules and ions being important modulators of biomolecular stability, dynamics and folding. It is now common to simulate biomolecules in a bulk solvent environment in which water is treated explicitly. However, modelling an aqueous

ionic environment is not trivial due to the fact that interactions between salt and salt, water and salt and water and water represent a balance of small additive van der Waals interactions, large electrostatic and dipole interactions, and large changes in entropy [123]. Numerous water models have been developed to predict the physical properties of liquid water [124]. Commonly used water models include the non-polarisable rigid models TIP3P [125], SPC/E [126], and TIP4E$_{EW}$ [127].

MD simulations in explicit solvent typically involve systems containing thousands of particles. Water molecules constitute over 80% of the particles and thus water-water interactions account for most of the computational cost. Reduction of the number of particles involved in the simulation can be achieved by imposing periodic boundary conditions on the system. Periodic boundary conditions permit the simulation of a relatively small number of particles of a large system as if they were in bulk fluid. Consider for example a cubic box of particles - the simulation cell - which is replicated in all directions thus forming a lattice of pseudo-infinite dimensions representing the bulk material. The primary simulation cell is thus surrounded by exact replicas (image cells) containing images of the particles in the simulation cell. Image particles in the replica cells move exactly as those in the simulation cell and periodic boundary conditions are implemented such that when particles move out from one side of the computational cell their images enter simultaneously from the opposite side. Pairwise interactions between particles $i$ and $j$ are calculated only between particle $i$ in the simulation cell and its closest particle $j$ be it the original particle or its image as shown in Figure 2.1. This, so-called *minimum image convention,* sets limitations on the interaction cutoff distance used and ensures that no duplicate interactions are computed. The cut-off distance, $r_{cutoff}$, defines a sphere of radius r, centered on particle $i$, and consistent with the minimum image convention the sphere should lie within the

simulation box. In general if the length of the side of the simulation box is $L$, $r_{cutoff} < L/2$.

While a cubic simulation cell represents the simplest geometry, periodic boundary conditions can be applied to other space-filling cells, such as the rhombic dodecahedron and the truncated octahedron. These space-filling models significantly reduce the number of water molecules within the simulation cell leading to a corresponding increase in computational efficiency [104]. However, the use of different computational cells implies that different minimum image distance criteria must be applied.

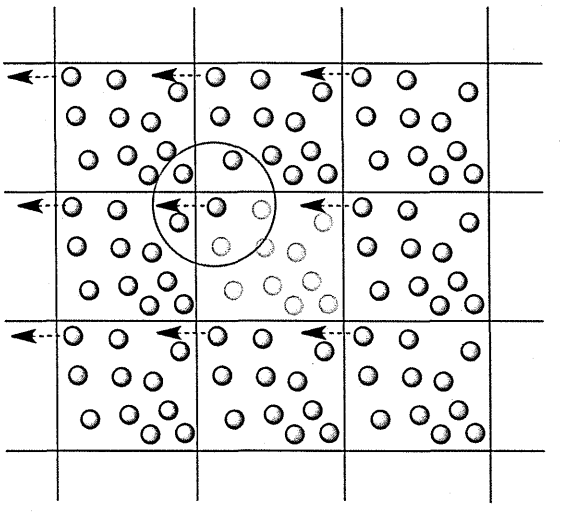


**Figure 2.1** **A two-dimensional slice of the periodic box, and the cutoff distance**

To calculate the force on an atom in the computational box, the pair potential of that atom with any other atom needs to be summed up. Differentiation of the potential with respect to the coordinates of the atom in question gives the force on that atom. The pair potential beyond the circumference of the red circle is negligible. Atoms in the central computational box are shown as light blue, and atoms in the replica boxes as grey

## 2.1.7 Modelling solvent effects

Biomolecular interactions normally occur in an aqueous environment and, consequently, solvent effects are particularly important determinants of protein-ligand binding. Such effects can be broadly grouped as electrostatic, hydrophobic

and pH-related. The interaction of each fully hydrated atom of a solvated molecule with its high-dielectric environment gives rise to what is known as one-body solvation energy for each atomic charge. The desolvation of polar moieties of a molecule brought about by conformational changes results in a substantial energy penalty. Charge-charge interactions of fully hydrated atoms are screened by water approximately 80-fold, i.e. the dielectric constant of water. On desolvation, these atoms interact with an effective dielectric constant ranging from 1 for fully desolvated atoms (i.e. the dielectric constant of a vacuum), to ≤ 80 for partially or fully solvated ones. Furthermore, the electrostatic energy of two atoms depends on both the interatomic distance and the positions of all other atoms of the solvated molecule, as the latter determine the solvent accessible surfaces of the molecule.

Hydrophobic effects are also all too important in biological processes. For example, the packing of non-polar amino acids, which is considered to be the most important contributing factor to protein stability, is driven by the so called "hydrophobic effect". Although the precise molecular basis of the hydrophobic effect is still the subject of some debate, the general consensus is that it is basically entropic in origin. Bulk water is highly structured through an extensive network of hydrogen bonded interactions. As non-polar groups can neither accept nor donate hydrogen bonds, any such group inserted into this structure disrupts the hydrogen bonding network of the water molecules. As a consequence of the formation of the cavity enclosing non-polar moieties, hydrogen bonding energy is lost. This loss is recovered to some extent by the re-orientation of water molecules at the interface with the cavity and the formation of hydrogen bonds with each other in such a way as to form a cage-like structure around the cavity. This locally ordered quasi-solid structure, also known as clathrate structure or "iceberg" results

in a loss of degrees of freedom of water molecules and therefore in a loss of entropy. The packing of non-polar groups minimises the area exposed to water as the cavity enclosing the aggregated non-polar groups is smaller than the sum of the surface areas of the cavities formed by the individual non-polar moieties. As a result, the conformational entropy loss of the entire system is minimised and the energy penalty in driving the hydrophobic moieties together in water is lowered. The change in the conformational entropy of water ($\Delta S_{solvent}$) is thus driving the aggregation of hydrophobic groups. The term "hydrophobic forces" has been coined to describe the effect of $\Delta S_{solvent}$.

Modelling the effects of the solvent on solute structure and dynamics has a direct bearing on the computational cost required to adequately explore the conformational space of biomolecules and the surrounding ionic environment. Computational methods for studying the dynamics of biomolecular systems in solution treat the solvent either explicitly, or implicitly.

#### 2.1.7.1 Explicit solvent methods

These methods represent the solvent in full atomic detail. Interactions between solvent, ions and solute are described by molecular mechanics force fields. These methods offer the full details of solvent-solute and solvent-solvent interactions but are very expensive computationally as they take into account the extra degrees of freedom associated the the explicit representation of the solvent and ions. For example, for a medium size protein of 350 amino acids, or approximately 5,500 atoms, immersed in a cubic box of 10-12 $Å^3$ there would be an additional 20,000-30,000 atoms involved in the simulation. Furthermore, the time of simulation would have to be increased considerably to obtain meaningful averages

of solute structure, dynamics or energetics from its instantaneous interactions with the explicit solvent [128].

### 2.1.7.2 Implicit or 'continuum' solvent methods

These methods are less accurate but have the advantage of being much less expensive in terms of computational cost. Implicit solvent methods, despite differences in implementation, treat the solute as an array of fixed atomic point charges embedded in a low dielectric continuum surrounded by solvent that is treated as a high dielectric continuum, and a diffuse cloud of charge representing aqueous ions. The solute-solvent interface is defined by the van der Waals envelope of the solute. Continuum solvent models capture non-specific solute-solvent interactions but do not account well for the energy contributions of different types of atoms and neglect the effects of ionisation of charged groups.

Implicit solvent methods reduce the degrees of freedom of the system by an 'averaging', that is by treating the molecules of interest explicitly and representing the solvent by its average effect on the solute [129]. Taking the example of a medium-size protein mentioned above would require sampling approximately 30,000 degrees of freedom. This is almost an order of magnitude less than the corresponding explicit water model. Implicit solvent methods split the solute-solvent interactions into a polar and an apolar component according to:

$$\Delta G_{solv} = G_{polar} + G_{npolar} \qquad 2.24$$

Solute-solvent interactions are described by solvation energies as shown in Figure 2.2. In this figure the solvation process is represented as a thermodynamic cycle consisting of three steps. In step 1, the solute is gradually stripped of charges in vacuum. The uncharged solute is then transferred from vacuum to solvent (step 2)

and an equal number of charges as those that were removed in step 1 are gradually added back to it (step 3). The sum of energies of steps 1 and 3 is the polar solvation free energy and represents the effect of solvent on the solute charging process, whereas the free energy change associated with step 2 is known as the non-polar solvation free energy.

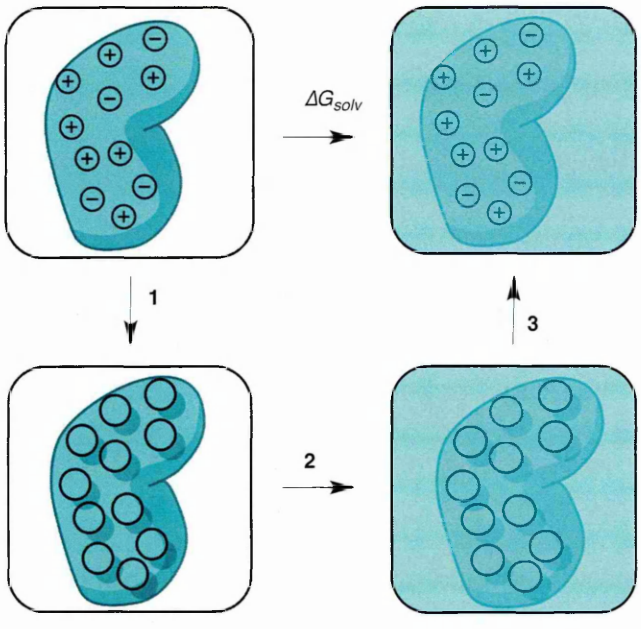



**Figure 2.2** **Biomolecular solvation: a thermodynamic cycle**

Step 1: Removal of charges *in vacuo*; step 2: uncharged biomolecule is transferred into solvent: step 3: charging the biomolecule in solvent to its normal charge value. The free energy change associated with step 2 is the nonpolar solvation free energy. The sum of free energy changes in steps 1 and 3 is the polar solvation free energy

These two free energy solvation terms act in opposite directions with polar solvation favouring maximum solvent exposure of all polar groups in the solute and non-polar solvation favouring compact solute structures of minimum solvent exposure.

### *2.1.7.2a Electrostatic contributions to the solvation free energy*

The Poisson-Boltzmann (PB) equation [105],[130], provides one of the most accurate continuum descriptions of the electrostatics of a biomolecular system. It

is a nonlinear differential equation that is usually approximated as a series expansion:

$$\nabla\varepsilon(\chi)\nabla\phi(\chi)+\varepsilon\kappa^2(\chi)\varphi(\chi)=-4\pi\rho(\chi) \qquad 2.25$$

where the first term of the equation is the Poisson equation which describes the electrostatic potential $\phi(\chi)$ at point $\chi$ generated by a charge distribution $\varrho(\chi)$ in an environment with dielectric permittivity coefficient $\varepsilon(\chi)$. The second term $\varkappa^2(\chi)$ is proportional to the ionic strength of the bulk solvent, that is the solvent outside the ion-accessible surface of the solute. The equation cannot be solved analytically for macromolecules of realistic geometry and charge distribution. Numerical solutions of the equation are available based on the discretisation of the domain of interest into small regions [130]. A particularly popular approach is that introduced by Warwicker and Watson [131]. This is a finite difference approach involving the superimposition of a cubic lattice over the biomolecular system and assigning values of electrostatic potential, charge, dielectric constant and ionic strength at each point of the lattice. The derivatives of the PB equation are then determined iteratively as the potential at each grid point influences the potential at neighbouring grid points.

Numerical solutions of the PB equation [130] give the electrostatic potential for the entire system from which it is possible to calculate the electrostatic free energy with a variety of different integral formulations the simplest one of which is:

$$G_{elec}^{PB}=\frac{1}{2}\sum_{i=1}^{N}Q_i\varphi(\chi_i) \qquad 2.26$$

where $Q_i$ is the point charge located at the centre of atom $i$.

Simpler implicit water models have been constructed the most popular of which is the generalised Born (GB) model. Generalised Born models are based on the analytical solutions of the equation derived by Born [132].

Equation 2.27, known as the Born equation, gives the electrostatic free energy of transfer of a spherical ion with charge $q$ and radius $a$ from vacuum (dielectric $\varepsilon_{vac}$) to a medium of dielectric $\varepsilon_{solv}$., and $k$ is the Coulomb constant, $(1/(4\pi\varepsilon_0))$, where $\varepsilon_0$ is the electrical permittivity.

$$\Delta G_{Born} = \frac{k^2}{2}\left(\frac{1}{\varepsilon_{solv}} - \frac{1}{\varepsilon_{vac}}\right)\frac{q^2}{a} \qquad 2.27$$

The Born model of solvation can be generalised for a system comprising of $N$ particles of arbitrary geometry, such as a protein, by treating each atom as a sphere of radius $a_i$, charge $q_i$. The total electrostatic energy of such system is given by:

$$G_{elec} = \sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{q_i q_j}{\varepsilon r_{ij}} - \frac{1}{2}\left(1 - \frac{1}{\varepsilon}\right)\sum_{i=1}^{N}\frac{q_i^2}{\alpha_i} \qquad 2.28$$

In the above equation, $r_{ij}$ is the interatomic distance, and the total electrostatic energy of the system is given as the sum of the Coulombic interactions (first term) and the Born free energy of solvation in a medium of relative permitivity $\varepsilon$ (second term). Unfortunately, for biomolecular systems the solution of Equation 2.28 is not possible as $a_i$ cannot be calculated with reasonable accuracy. For a simple solute, such as a metal ion, $a$ can be considered to be equal to the van der Waals radius of the solute, or alternatively equal to the distance from the atomic charge to the dielectric boundary. However, for more complex systems, the radius $a_i$ of atom $i$ depends on the volumes and positions of all other atoms of the solute.

Still [133] has shown that the first term of Equation 2.28 can be expanded algebraically to give the Coulomb energy *in vacuo* and a second term that accounts for the effect of the dielectric medium on the pairwise interactions of the charged particles.

$$G_{elec} = \sum_{i=1}^{N}\sum_{j\neq i}^{N}\frac{q_i q_j}{r_{ij}} - \left(1-\frac{1}{\varepsilon}\right)\sum_{i}^{N}\sum_{j\neq i}^{N}\frac{q_i q_j}{r_{ij}} - \frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{i=1}^{N}\frac{q_i^2}{a_i} \qquad 2.29$$

The generalised Born equation (GB) is the difference between the total electrostatic energy of the system, $G_{elec}$, and the Coulomb energy *in vacuo*, i.e. the first term of Equation 2.29:

$$\Delta G_{elec}^{GB} = -\frac{1}{2}\left(1-\frac{1}{\varepsilon}\right)\sum_{i}\sum_{j}\frac{q_i q_j}{\sqrt{r_{ij}^2 + B_i B_j e^{\frac{-r_{ij}^2}{4B_i B_j}}}} \qquad 2.30$$

where $B_i$, $B_j$ are the "effective Born radii" by analogy to $a_i$ in Equation 2.28, and $r_{ij}$ are the effective interaction distances between two atoms. $B_i$ is defined as the radius that would give the self-energy of atom $i$ of a biomolecular system if all other atoms in the system were uncharged thus only displacing the dielectric. According to this definition, $B_i$ would mark out a spherically averaged dielectric boundary around atom $a_i$.

In principle, such "perfect" radii could be derived by solving the Poisson equation but this would nullify the computational advantage of the GB model. Still [133] and co-workers proposed a numerical finite difference method for a reasonably accurate calculation of Born radii. A less accurate analytical calculation of Born radii of lower computational cost has also been developed. Both types of

calculation used in conjunction with the GB water model reproduce solvation energies in good agreement with experimental results [134].

#### *2.1.7.2b Non-electrostatic contributions to the solvation free energy*

The generalised Born (GB) and Poisson-Boltzmann (PB) electrostatic models have been enhanced with the addition of a surface area-related (SA) non-polar solvation energy term. These enhanced models are known as the GBSA and PBSA solvation models, respectively. The non-polar solvation term, which is-the most heuristic component of continuum solvation models [128] consists of two terms, one related to the formation of a cavity upon immersion of the protein into the solvent (Section 2.1.7), and the second related to attractive van der Waals interactions between the protein and the solvent.

In most continuum solvation models, the non-polar component of the solvation free energy is taken to be linearly dependent on the solvent-accessible surface area (SASA) of the macromolecule:

$$G_{solv}^{npolar} = G_{solv}^{cavity} + G_{solv}^{vdW} = \gamma \sum_{i=1}^{N} A_i \qquad 2.31$$

where $A$ is the surface area defining the region around atom $i$ that is inaccessible to the solvent. The coefficient $\gamma$, which are also known as "surface tension" parameters, are empirically derived from sets of coefficients used to reproduce solvation free energies of alkanes.

Chothia has demonstrated that the non-polar solvation free energy is linearly dependent on SASA [135]. The "accessible surface" is defined as the area circumscribed by the centre of a spherical probe (usually a water molecule) rolling over the van der Waals envelope of a protein (or of a side chain). Several methods

exist to calculate SASA most of which are based on the Shrake-Rupley algorithm [136]. SASA can be considered as a proxy representation of the hydrophobic effect discussed in Section 2.1.7.

SASA models are subject to caveats, such as the actual choice of "surface tension" parameters, as well as the accessibility of buried atoms to the solvent. Some of these problems have been addressed by more complete models which account for attractive van der Waals interactions between solute and solvent, as well as repulsive solvent accessible volume terms [137], [138].

### 2.1.8 Langevin Dynamics

Commonly, biological systems are studied under constant temperature conditions as changes in volume and volume energies are insignificant are in most cases not significant. Constant temperature conditions can be attained by bringing the system in thermal contact with a heat bath. Temperature control is achieved by means of one of several types of thermostat, examples of which are the Berendsen, Langevin and Nosé-Hoover thermostats. These thermostats control temperature in different ways. For example, whereas the Berendsen thermostat [139] corrects deviations from the desired temperature by multiplying the velocities by a factor, the Nosé-Hoover thermostat [140] introduces additional degrees of freedom (momentum) into the Hamiltonian of the system.

The Langevin thermostats follow the Langevin equation of motion as an alternative to Newton's second law [141]. This equation adds two more terms into Newton's equations of motion the first of which is a damping function representing the frictional drag experienced by the solute. The second additional term is a fluctuating force that is applied to represent the random kicks from the thermal motion of the solvent molecules:

$$m_i\vec{a} = \vec{F}_i(r) - \zeta_i\vec{v} + \vec{R}_i(t) \qquad 2.32$$

where $m_i$ is the mass of particle $i$, $a$ is the acceleration, $F_i(r)$ is the term used in conventional MD, $R_i$ the random force exerted on particle $i$ and $\zeta$ the frictional coefficient. The frictional force $\zeta v$ has a fixed positive value and decreases the temperature. $R_i$ is randomly determined from a Gaussian distribution, and its variance is the function of the set temperature and time step. It adds kinetic energy to the particle to counterbalance the frictional force and maintain the system temperature at the set value. When $\zeta_i = 0$ Langevin dynamics reduces to conventional MD, whereas when $\zeta_i > 0$ the random impulses on the system brought about by $R_i$ may induce energy barrier-crossing motions and thus could lead to improved conformational sampling characteristics over conventional MD.

### 2.1.9 Steered Molecular Dynamics (SMD)

The mechanism of macromolecular complex association and dissociation is of critical functional importance in many biological processes such as, for example, the recognition of ligands by proteins. However, such processes involve rare barrier-crossing events, which are too slow to reproduce in the timescales of current MD simulations. The problem can be addressed by applying external forces to guide the system from one state to another thus accelerating processes which are difficult to model.

Computer simulations describing association and dissociation events between protein and ligand strive for reversibility, as for example, in umbrella sampling and free energy perturbation [142]. Attaining ideally reversible processes requires extremely slow manipulation and consequently a very high computation cost. The application of an external force on the system, on the other hand, accepts

irreversibility in order to gain access to biologically relevant information related to protein-ligand binding and dissociation. The question is how to obtain thermodynamic potentials from such irreversible processes. The answer to this is by the so-called Jarzynski relationship a formalism of which is given by:

$$\exp(-\Delta G / k_B T) = \langle \exp(-W / k_B T) \rangle \qquad 2.33$$

where the brackets $< >$ denote averages, $k_B$ is the Boltzmann constant, $T$ the temperature, and $W$ work done through a non-equilibrium process. Thus computing the work done during a non-equilibrium transition from one state to another and averaging over the initial state allows the calculation of equilibrium free energies from non-equilibrium calculations [143].

In SMD this is achieved by applying an external force onto a system to drive a change of coordinates within a given time. In summary, if a system is subject to an external time-dependent perturbation ($\lambda(r) = \lambda(t)$) defined by a harmonic potential, its Hamiltonian, $H(r,\lambda)$, can be written as the sum of the time independent Hamiltonian of the unperturbed system, $H_0(r)$, plus the time-dependent external potential. Assuming that a system is steered from its original state A to state B, the Hamiltonian of the system is a function of $\lambda$ and represents the system in the state A when $\lambda = 0$ ($\lambda_0$) and in the state B when $\lambda \neq \lambda_0$. The minimum position of the harmonic potential moves along a chosen reaction path $\lambda(r)$ at constant velocity $(v)$ according to:

$$H(r,\lambda) = H_0(r) + (1/2)k[\lambda(r) - \lambda_0 vt]^2 \qquad 2.34$$

where $\lambda(r)$ represents the chosen reaction path, and $r$ represents the configuration of the whole system. The free energy of a process can thus be obtained by

performing a finite number of transformations, calculating the work done at each step, and averaging as in Equation 2.33.

### 2.1.10 Binding free energy calculations

As mentioned earlier (Section 2.1.2), MD simulations are often performed to elucidate the thermodynamic properties of model systems. Free energy calculations in particular are routinely used to evaluate relative stabilities of macromolecular structures, and to predict protein-ligand binding affinities. A number of methods based on MD have been devised to calculate free energy and predict binding affinity, including thermodynamic integration [100], single-step perturbation [144], and free-energy perturbation [99]. These methods are computationally intensive, due to their treating the solvent explicitly. Alternative less rigorous techniques of lower computational cost have been developed, which nevertheless have been proven to be remarkably accurate in predicting binding affinities [145]. One such method, the Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) and its variant the Molecular Mechanics Generalised Born Surface Area (MM-GBSA), are presented in some detail below.

The MM-PB(GB)SA methods are basically MD post-processing methods. They are applied subsequent to MD simulations, which are usually performed in explicit solvent to generate conformational ensembles. Explicit solvent and ions are stripped from the trajectory files to prevent solvent-solvent interactions from dominating the energy terms, ensemble structures are then extracted and free energies calculated. Free energies obtained by these methods can be decomposed into a molecular mechanics energies, continuum solvation energies and solute entropy terms according to:

$$\langle G \rangle = \langle E_{MM} \rangle + \langle G_{PBSA} \rangle - T \langle S_{MM} \rangle \qquad 2.35$$

where $\langle G \rangle$ is the average free energy of the solute (protein, ligand or complex), $\langle E_{MM} \rangle$ the statistical average of equation 2.8, $\langle G_{PBSA} \rangle$ is the average of the sums of Equations 2.26 and 2.31, $T$ is the temperature and the average $\langle S_{MM} \rangle$ represents the entropic contribution of the solute. $S_{MM}$ can be decomposed into entropic contributions made by rotational, translational and vibrational motions of the solute. It is usually calculated by applying either normal mode analysis [146] or quasi-harmonic analysis [147].

Both MM-PBSA and the computationally more efficient MM-GBSA variant of the method can be used to study the relative stability of different conformations of the same macromolecule. This is done by running independent simulations on different conformers of a protein, post-processing the trajectories by stripping the solvent and comparing the free energies of the conformers, as shown in Figure 2.3.

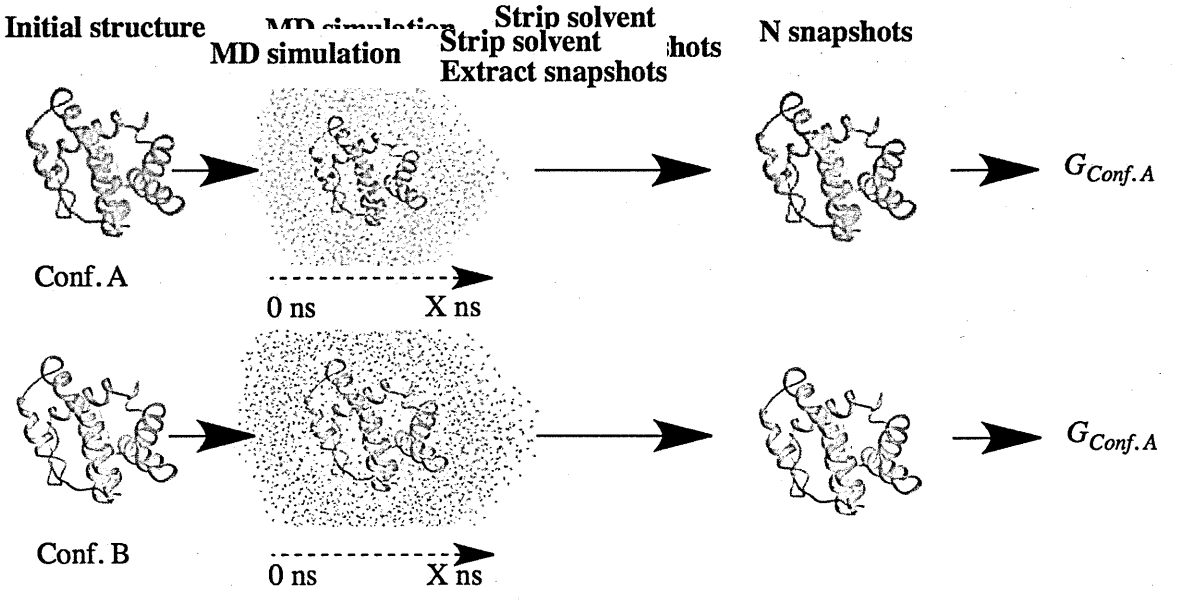



**Figure 2.3** **Comparison of the stability of two protein conformations**

Another application of the MM-PB(GB)SA methods is to evaluate the binding free energies of protein-ligand complexes.

The non-covalent association of a ligand and a receptor molecule is described by:

$$[receptor]_{solv} + [ligand]_{solv} \rightleftharpoons [complex]_{solv} \qquad 2.36$$

The formation and stability of the complex does not only depend on the binding energy but also on the Gibbs free energy of the system. The term "free energy" is used for "Gibbs free energy" and the term "binding energy" for "binding enthalpy". In most biological systems changes in volume and volume energies are insignificant and therefore Gibbs free energy equals Helmholtz free energy and binding energy equals binding enthalpy. The free energy of binding can be computed according to:

$$\Delta G^{solv}_{binding} = G^{solv}_{complex} - G^{solv}_{receptor} - G^{solv}_{ligand} \qquad 2.37$$

where $G$ refers to the solvation free energy of each of the three molecular species [148].

MM-PB(GB)SA employs the thermodynamic cycle shown in Figure 2.4. and computes absolute binding free energy ($\Delta G_{binding}$) according to Equation 2.37. As only the initial and final states of the system are evaluated to estimate free energy changes, the method is known as *end-point free energy* method. $\Delta G_{binding}$ is estimated as the difference between the free energies of a complex and the sum of the free energies of its components.

$$\Delta G_{binding} = \Delta E_{MM} + \left(\Delta G^{complex}_{solv} - \Delta G^{protein}_{solv} - \Delta G^{ligand}_{solv}\right) - T\Delta S_{config} \qquad 2.38$$

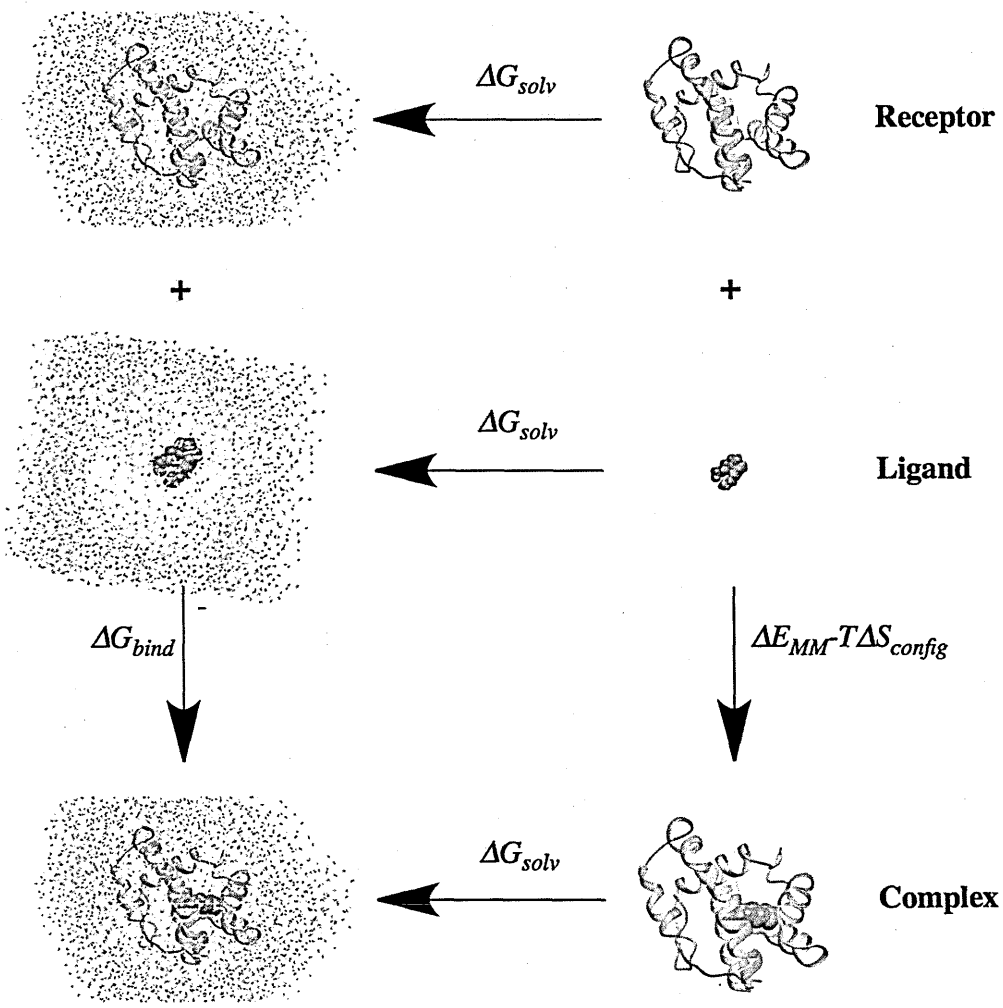



**Figure 2.4** **Thermodynamic cycle: Calculation of a binding free energy**

$\Delta G_{solv}$ is the change in the free energy of solvation for each molecular species

MM-PB(GB)SA can be used to calculate $\Delta G_{solv}$ either by, what is known, as the "single-trajectory" or the "three-trajectory" approach (Figure 2.5). In the three-trajectory approach, independent MD simulations are performed for each molecular species *(i)*, the trajectories are post-processed as described above, and $\Delta G_i$ is evaluated. $\Delta G_{binding}$ is then calculated according to equation 2.38.

In the single-trajectory approach, a single MD simulation of the complex is performed, and snapshots are sampled from the post-processed trajectory. The method assumes that the configurational space explored during an MD simulation is very similar between the bound and unbound states and thus snapshots for each molecular species can be extracted from the same trajectory file. Strictly

speaking, this is not true but the method is computationally less costly than the three-trajectory approach and the results obtained have been found to be in good agreement with experiment.

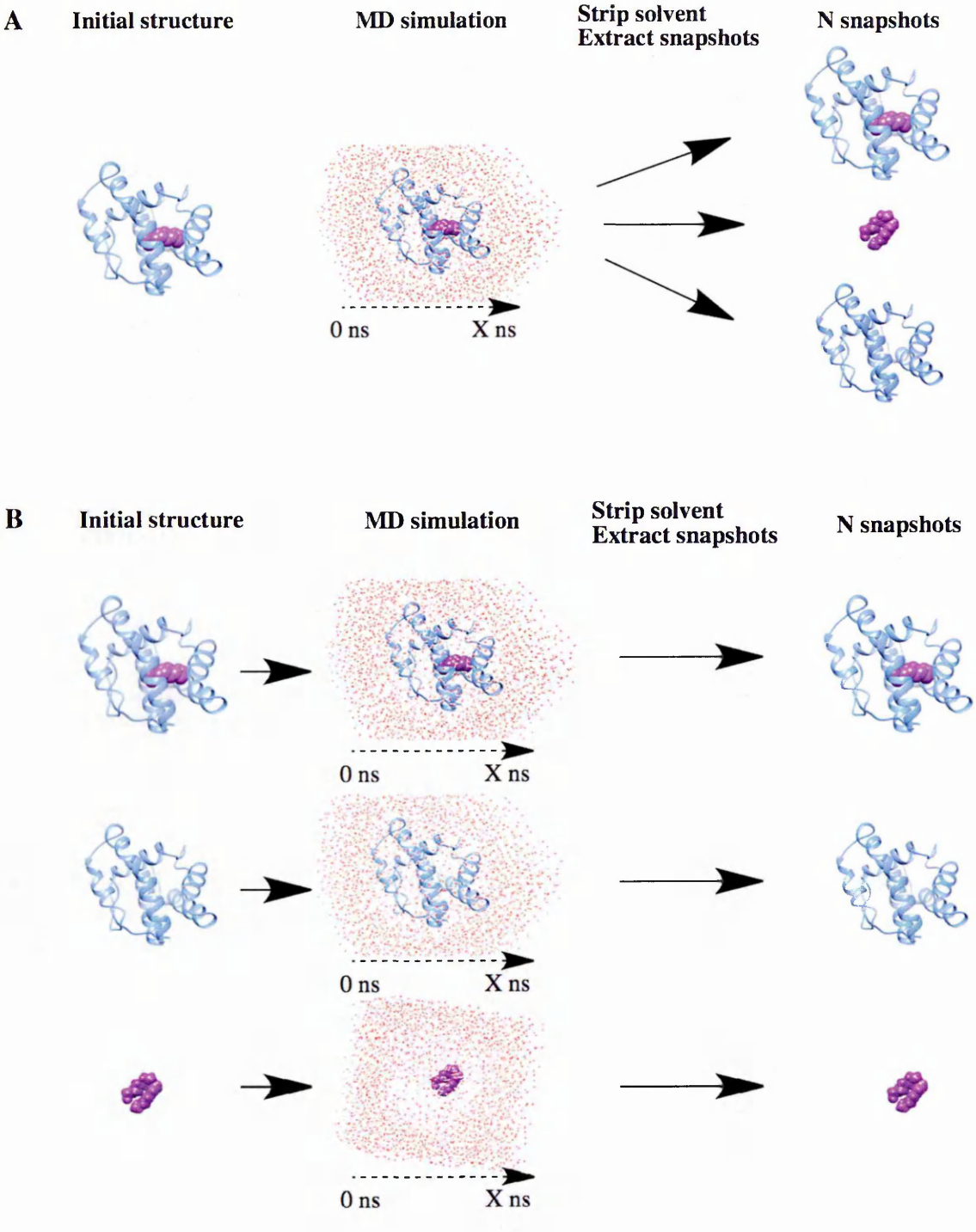



**Figure 2.5** Binding free energy calculations: single- (A) and three-trajectory (B) approach

### 2.1.11 Docking

Molecular docking is another powerful technique used in structure-based drug design where it has two main practical applications. One of these, known as blind docking, aims at determining the most probable docking pose of a ligand in the binding pocket of its receptor when appropriate crystallographic or NMR models are not available. The second, known as virtual screening, involves the binding of of ligand libraries to a particular receptor. Blind docking searches the entire protein surface for binding sites while optimising the conformations of the molecular entities involved. In virtual screening the main objective is not the determination of the most probable ligand poses, but instead, the relative binding affinities of the docked ligands for the protein receptor. Docking techniques achieve a reduction of computational expense by utilising simpler molecular mechanics force fields and solvent models combined with empirical weights and/or empirical functional forms.

Essentially molecular docking works by evoking two basic methods: (i) a search method for exploring the conformational landscape of the system and (ii) a force field to evaluate the energetics of the sampled conformations. Modern docking methods are able to predict the correct ligand binding mode in over 70% of cases [149] and the most stable protein/ligand complex derived from a docking simulation is used to assign to the ligand a binding energy or score. This can be done either by using the same force field as that of the docking simulation or by using a more sophisticated force field.

Scoring functions can be grouped into three main categories: empirical, knowledge-based and physical (based on a molecular mechanics force field). An example of an empirical scoring function is GOLD [150], which consists of a hydrogen binding, a van der Waals and an internal energy term. DrugScore$^{CSD}$ is

an example of a knowledge-based scoring function, which uses a formalism derived from potentials of mean force [151]. Lastly, an example of a physical scoring function is AutoDock, which uses a semi-empirical force field [152].

Typically, the scoring functions discussed above are able to discriminate between good ligand binders and poor ones but fail to rank a series of ligands according to binding affinities [153]. This is not surprising as binding affinity is determined to a large extent by fine hydrogen bonding interactions, and hydrophobic contacts between the ligand and the receptor. Moreover, the evaluation of binding free energy is dependent on solvation and entropic factors, which are often not considered in most common scoring functions.

## 2.2 Methods

### 2.2.1 AutoDock

The interaction of putative odorant molecules (ligands) with PBP/OBP targets was studied by docking simulations using AutoDock 4.2 [152] and AutoDock Vina [154]. The semi-empirical force field employed by AutoDock4 uses pair-wise terms to evaluate the interaction energies between two or more molecules in an aqueous environment and an empirical method to determine the contribution of water. The difference between this approach and that of a traditional molecular mechanics one is that the latter can use explicit water molecules to evaluate the contribution of solvation/desolvation to the binding energy. AutoDock's force field evaluates protein-ligand binding in two steps as shown in Figure 2.6. In the first of these steps, the intramolecular energy difference for the transition from the unbound to bound conformations of each of the molecular species is evaluated separately. The second step evaluates the intermolecular energetics of complexation by bringing the two molecules together.

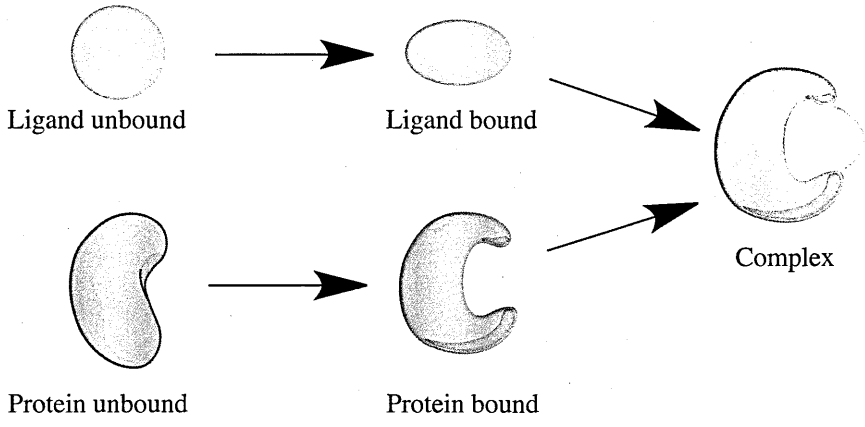



**Figure 2.6** **Autodock: Binding evaluation**

The Autodock force field evaluates binding in two steps. Step 1: Ligand and protein start in an unbound conformation,and the intermolecular energetics are estimated for the transition to the conformation of the ligand and protein in the bound state. Step 2: The intermolecular energetics of complex are computed by combining the ligand and protein in their bound conformation

The force field has been developed from data derived from a large number of studies involving protein-inhibitor complexes and inhibition constants of known structures, and is applied in two stages. In stage one, the force field is used to estimate the intramolecular energetics for the transition of the unbound conformations of the ligand and protein to the bound state. During stage two, it is applied to evaluate the intermolecular energetics of the ligand and protein in their bound state. The resulting change in the free energy of the system upon binding is computed from six pair-wise evaluations ($V$) as well as an estimate of the conformational entropy lost upon binding $\Delta S_{conf}$:

$$\Delta G = \left(V_{bound}^{L-L} - V_{unbound}^{L-L}\right) + \left(V_{bound}^{P-P} - V_{unbound}^{P-P}\right) + \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + T\Delta S_{conf}\right) \qquad 2.39$$

where $L$ is the "ligand" and $P$ the "protein".

Each evaluation is made up by computing (a) the dispersion/repulsion interactions described by a 6/12 potential, using empirical parameters $A$ and $B$ derived from the Amber force field; (b) a 10/12 potential measuring directional H-bond formation; (c) Coulomb potential electrostatics, and (d) the desolvation potential based on the

volume of atoms ($V$) that surround a given atom sheltering it from the solvent according to:

$$V = W_{vdw}\sum_{i,j}\left(\frac{A_{ij}}{r_{ij}^{12}} - \frac{B_{ij}}{r_{ij}^{6}}\right) + W_{hbond}\sum_{i,j}E(t)\left(\frac{C_{ij}}{r_{ij}^{12}} - \frac{D_{ij}}{r_{ij}^{10}}\right) + W_{elec}\sum_{i,j}\frac{q_i q_j}{\varepsilon(r_{ij})r_{ij}} + W_{sol}\sum_{i,j}\left(S_i V_j + S_j V_i\right)e^{\left(-\frac{r_{ij}^2}{2\sigma^2}\right)}$$

2.40

The weighting constants $W$ have been optimised from a set of experimentally-determined binding constants. The parameters $C$ and $D$ give a maximal energy well depth of 5 kcal/mol at 1.9 Å for hydrogen bonds with oxygen and nitrogen, and a well depth of 1 kcal/mol at 2.5 Å for hydrogen bonds with sulfur. $E(t)$ is a function providing the directionality of H-bond formation at an angle $t$ from ideal h-bonding geometry. The solvation term of the equation is weighted by parameter $S$ and a distance weighting factor $\sigma$ = 3.5 Å. The term for the loss of entropy upon binding is directly proportional to the number of rotatable bonds.

The way AutoDock converts binding energies into binding affinities ($K_b$) or dissociation constants ($K_d$) is based on the simple equation:

$$K_b = \frac{[L][P]}{[PL]} \qquad 2.41$$

where the brackets denote an equilibrium concentration, $L$ and $P$ are the ligand and protein, respectively, and $PL$ the ligand-protein complex. An inherent assumption of this equation is that the bound and unbound states are clearly distinguishable. This assumption is essentially correct for tight binders but not so for very weak or non-specific binders.

$K_b$ and $K_d$ are related by:

$$K_b = 1/K_d \quad \text{or} \quad \ln K_b = -\ln K_d \qquad 2.42$$

The dissociation constant is therefore related to the free energy of binding by:

$$K_d = \exp^{\Delta G_{binding}/RT} \qquad 2.43$$

The non-covalent protein-ligand association to form a complex is therefore an equilibrium property relating the free energy of binding ($\Delta G_{binding}$) and the equilibrium constant ($K_b$) logarithmically. This means that differences in the binding free energy of only 1.4 kcal/mol result in a 10-fold change in binding affinities.

## 2.2.2 Docking simulations

### 2.2.2.1 Re-docking and cross-docking

Re-docking and cross-docking simulations were conducted using the Lamarckian genetic search algorithm (LGA) and the semiempirical force field of the AutoDock (v.4.2) program [155]. The AutoDock Tools [156] GUI was used to remove crystallographic waters, with the exception of water molecules buried in the binding pocket, and to add polar hydrogens to proteins and ligands. The MMFF94 force field [157] was used for energy minimisation of the ligands. Ligand preparation involved the release of all torsions except those around conjugated double and triple bonds and Gasteiger partial charges were added while non-polar hydrogen atoms were merged. Protein preparation involved the addition of essential hydrogen atoms, Kollman united atom charges and solvation parameters. Proteins were covered, in their entirety, by affinity (grid) maps with spacing of 0.374 Å. The default parameters of AutoDock were used for distance-dependent dielectric functions, van der Waals and electrostatic terms. The starting positions and orientation of the ligands were set randomly. Ligands were subjected to 100-200 LGA runs with a number of evaluations ranging between (0.5 - 25) $10^6$ depending on the number of torsions in each ligand. The default AutoDock force

field was applied [152]. The RMSD tolerance of the resulting docked structures was ≤ 2Å. Docking results were sorted into bins of similar conformations. Cluster analysis or 'structure binning" was performed based on all-atom mean square deviation (RMSD). The resulting families of docked conformations were ranked in order of increasing energy (rank 1 was taken to be the lowest energy cluster).

#### 2.2.2.2 Docking of ligand libraries

Virtual screening of ligand libraries was performed with AutoDock Vina [154].The ligands were prepared for docking with the PyRx virtual screening software [158]. Ligand preparation involved minimisation using the MMFF94 force field [157] and the addition of Gasteiger partial charges. Proteins were prepared as described in Section 2.2.2.1 the only difference being the spacing of the affinity (grid) maps, which was 1 Å instead of 0.374 Å that was used in docking simulations performed with AutoDock. Ligand efficiency (*LE*) and fit quality (*FQ*) scores [159] were used to obtain relative rankings of binding affinities of the docked ligands. Ligand efficiency (*LE*) is any property defining binding affinity, such as *pIC50*, *pK*i, $\Delta G$, divided by the molecular size of the ligand. Ligand efficiency is also dependent on molecular size and reaches a plateau for ligands above 45 non-hydrogen atoms. Fit quality (*FQ*) is an alternative metric that normalises ligand efficiency with respect to the size of the ligand.

### 2.2.3 AMBER

Energy minimisations and MD simulations were carried out using the *sander* module of the AMBER suite of programs, which implements a force field of the following form:

$$V(r) = \sum_{bonds} k_l\left(l - l_0\right)^2 + \sum_{angles} k_\theta\left(\theta - \theta_0\right)^2 + \sum_{dihedrals} \left(V_n / 2\right)\left(1 + \cos[n\omega - \gamma]\right) + \sum_{nonbij} \left(A_{ij} / R_{ij}^{12}\right) - \left(B_{ij} / R_{ij}^{6}\right) + \left(q_i q_j / R_{ij}\right) \quad 2.44$$

The terms of the above equation are defined in Section 2.1.4.

Long-range electrostatic interactions are handled by the particle-mesh Ewald procedure (PME) [115, 116] and long-range van der Waals interactions are handled by a continuum model. Non-periodic simulations with aqueous solvation effects represented implicitly by GBSA/PBSA models can be carried out by adding the two terms of the r.h.s of the equation below to equation 2.44 [160].

$$\Delta G_{sol} = \sum_{ij}(1-1/\varepsilon)\left(q_i q_j / f_{GB/PB}\left(R_{ij}\right)\right)+\gamma\sum_i A_i \qquad 2.45$$

The first term of the above equation, which is designed to provide an approximation of the reaction field potential, accounts for the electrostatic contribution to the solvation free energy (see Section 2.1.7.2a). The function $f_{GB/PB}$ represents numerical solutions of the Poisson Boltzmann (PB) or the generalised Born (GB) equation, examples of which are given in Equations 2.26 and 2.30, respectively. The second term of the equation represents the non-electrostatic contributions to the solvation free energy as described in Section 2.1.7.2b.

### 2.2.3.1 MD equilibration and production simulations

The AMBER 11 and AmberTools 1.5 [160] programs were used for MD simulations and the analysis of the MD simulation data, respectively. Simulations were carried out in the isothermal isobaric thermodynamic ensemble at 300K using the ff99SB [161] and gaff [162] force-field parameters for the proteins and ligands respectively. The protein parameter and coordinate files were prepared using the LEAP module of AmberTools, whereas the corresponding ligand files were prepared using the Antechamber suite of AmberTools. All complexes were neutralised with the addition of the requisite number of $Na^+$ counter-ions using the

LEAP module. Each system was immersed into a truncated octahedron periodic box containing water molecules. The TIP3P water model was used [125]. The number of water molecules used as ~ 4,000[3]. Periodic box boundaries were set at a distance of 9 Å from any solute atom. The systems were minimised by 500 steps of steepest descent minimisation followed by 500 steps of conjugate gradient minimisation keeping the solute fixed. The harmonic restraint on the solute atoms was 2 kcal $mol^{-1}$ $Å^{-2}$. Restraints on solutes were removed and a second round of 1000 steps of steepest decent minimisation followed by 1500 steps of conjugate gradient minimisation was performed.

Bond lengths involving bonds to hydrogen atoms were constrained using the SHAKE algorithm [163]. The particle mesh Ewald (PME) method with an 8 Å cutoff was used for the treatment of long-range electrostatic interactions. The time-step for all MD simulations was 2 fs. No smoothing or switching function were used [4]. This is because with PME, the full electrostatics are represented. Furthermore, there was no smooth cutoff to the van der Waals, as it is a hard cutoff. Although one can argue that other smoothing or switching of forces or energies may be more accurate, if the cutoff is in the 8-10 Å range, the effects are fairly minimal [164].

The systems were heated from 0 K to 300 K by carrying out a 50 ps canonical ensemble (NVT)-MD during which harmonic restraints were applied to all solute atoms with force constants of 0.1 kcal $mol^{-1}$ $Å^{-2}$. Subsequently, the systems were subjected to 50 ps of isothermal isobaric (NPT)-MD at 300 K with coupling to Langevin thermostat. Constant pressure periodic boundaries conditions were

---

[3] Exact figures are given in Table 3.9, p.116, and Table 4.4 p.168

[4] "The current cutoff scheme for non-bonded interactions in AMBER modules and NAB does not use a switching function to smooth the cutoff." Amber 13 manual, p.168

applied with (a) isotropic scaling and 2.0 ps relaxation time to maintain the pressure at an average of 1 atm [5], and (b) weak restraints on the solute atoms (0.1 kcal $mol^{-1}Å^{-2}$ ).

Finally, the systems were equilibrated for an additional 2ns NVT-MD simulation at 300 K with a time constant of 2.0 ps for Langevin bath coupling and removal of the constraints on the solute.

Equilibration of the system was followed by production time MD for each of the complexes over 10 ns unless otherwise stated.

### 2.2.4 SMD

The theoretical underpinnings of the method were presented in Section 2.1.9. . In the Amber implementation of SMD [160], for a group of atoms harmonically restrained with a force constant $k$, the additional potential is defined by:

$$V_{rest}(t) = \frac{1}{2}k[x - x_0(t)]^2 \qquad 2.46$$

where $x$ is a distance, an angle or a torsion, and $x_0$ is the 'center' of the restraint that is the value at which the restraint potential is zero. In this work, the external force was applied at constant velocity and the coordinate in question was distance. By integrating the force ($V_{rest}$) over time or distance a generalised work can be computed and from this free energy differences using the so-called Jarzynski relationship (see Section 2.1.9).

[5] Constant pressure is used to allow the density of the water to relax.The barostat is set up to rescale the volume isotropically, that is, it uses a linear transformation to correct changes in volume

### 2.2.5 Free energy calculations

Molecular Mechanics Generalised Born Surface Area calculations were performed by invoking the Amber MMPBSA program. The theoretical underpinnings of the method are given in Section 2.1.10. Binding free energies were calculated as averages of results obtained on a per-frame (snapshot) basis according to:

$$\Delta G_{total} = \langle G^{complex}(i) \rangle - \langle G^{protein}(i) \rangle - \langle G^{ligand}(i) \rangle \qquad 2.47$$

where $\langle\ \rangle$ is the average over trajectory snapshots $i$ and $G^x(i)$ is computed from gas-phase energy, solvation free energy and entropy contributions of each molecular species, $x$, according to:

$$G^x(i) = H^x_{gas}(i) + H^x_{trans/rot}(i) + G^x_{solvation}(i) - TS^x(i) \qquad 2.48$$

For each snapshot, the gas phase contributions ($H_{gas}$) were calculated by summing internal energy contributions (e.g. bond, angle, torsional angle energies). $H_{trans/rot}$ at 300K and in the classical limit equals 1.8 kcal/mol corresponding to the six translational and rotational degrees of freedom, according to:

$$H_{rot/trans} = 3RT = 1.8 kcal mol^{-1} \qquad 2.49$$

Solvation free energies were computed as the sum of polar and non-polar contributions. Non-polar contributions were approximated by the LCPO method [165] implemented within the Amber's *sander* program[6]. The electrostatic portion was calculated using both the PB and GB models described in Section 2.1.7. In PB calculations the internal and external dielectric constants were left at the default values, 1.0 and 80.0, respectively. The second modification of the Bondi radii set [166] was used for GB calculations.

[6] The $\gamma$ default value of 0.00542 kcal mol$^{-1}$ Å$^{-2}$ was used

Entropy calculations, which include the translational, rotational and vibrational entropy of the solute were calculated by means of gas phase statistical mechanics by the nmode program in Amber. The program calculates vibrational modes using harmonic approximation after energy minimisation in distance-dependent dielectric environment.

For the free energy calculations, snapshots were taken every 40 ps from the last 10 ns of production runs giving 250 sample points. The number of sample was deemed to be sufficient to calculate mean values with reasonable precision since the error in the mean is inversely proportional to the square-root of the number of snapshots [167]. This assumes that the sampled values are independent. It is not unreasonable to expect that since the sample interval time between calculations (40 ps) is much longer than the time for relaxation of effective energy fluctuations (1ps), the sampled values should be uncorrelated. As entropy calculations bear a high computational cost, they were performed only every 400 ps.

#### 2.2.5.1 Per-residue free energy decomposition

In addition to decomposing binding free energies into gas-phase energy, solvation free energy and entropy components, it is possible to decompose $\Delta G_{total}$ in terms of contributions arising from the interactions of the binding partners. This type of decomposition provides useful information regarding the origin of binding at the atomic level. One such method of per-residue free energy decomposition is alanine scanning [167]. An alternative method has been employed in this work which is thought to avoid some of the pitfalls of alanine scanning, namely the fact that even virtual mutations can perturb specific localised interactions at the binding interface. Considering the contributions of each residue to the binding free energy by means of component analysis, Equation 2.47 yields [168]:

$$\Delta G_{total} = \sum_{\substack{j \in complex \\ \wedge\, j \in protein}} \left( \left\langle G^{complex}(i,j) \right\rangle - \left\langle G^{protein}(i,j) \right\rangle \right) + \sum_{\substack{j \in complex \\ \wedge\, j \in ligand}} \left( \left\langle G^{complex}(i,j) \right\rangle - \left\langle G^{ligand}(i,j) \right\rangle \right) - H_{trans/rot} \qquad 2.50$$

where $\langle\ \rangle$ denotes averaging over all sampled snapshots $i$ and each of the addends provides the contribution to the binding free energy by residues $j$. The ligand is considered be a single residue. Considering that Equation 2.50 includes the energy contribution due to translational and rotational degrees of freedom, then according to Equation 2.48, $G^x(i,j)$ contains contributions from internal gas-phase energies, solvation free energies, and entropies:

$$G^x(i,j) = H^x_{gas}(i,j) + G^x_{solv}(i,j) - TS^x(i,j) \qquad 2.51$$

## 2.2.6 Trajectory analysis

MD trajectory analysis was performed with the AmberTools' *ptraj* and *cpptraj* suite of programs [169]. These programs process and analyse a series of atomic coordinates, one molecular configuration (frame) at a time. Molecular information such as residue and atom names is loaded from the Amber topology files (prmtop).

### 2.2.6.1 Trajectory clustering

*ptraj* was used to group together coordinate frames from trajectories into distinct sets. Thprogram provides an implementation of the average linkage clustering algorithm. Under this algorithm, the distance from one cluster to another is defined as the average of all distances between individual points of the two clusters [170]. The mass-weighed RMSD of the backbone C-atoms was used as a distance metric. The Davies-Bouldin (DBI) and the pseudo F-statistic (pSF) were used as metrics of the clustering quality. DBI is related to the inter-cluster scatter of any given cluster, as well as to the inter-cluster separation and is therefore a useful

measure of cluster compactness and separation. Low values of DBI are indicative of better clustering. pSF is a comparator of inter-cluster variance to the residual variance over all sampled points. High values of pSF are indicative of better clustering.

#### 2.2.6.2 Hydrogen bond and salt-bridge formation

Monitoring of hydrogen bond formation between ligands and proteins was performed using *cpptraj*. The default distance and angle cutoffs of 3.0 Å and 135°, respectively, were used. H-bond occupancies were recorded. Occupancy is defined as the fraction of simulation time that the hydrogen bond is formed. H-bond formation was also monitored using the H-bond plugin of VMD (Visual Molecular Dynamics) [171]. The default distance (3.0 Å) and angle (20°) cutoffs were used. Salt-bridge formation was monitored with the salt-bridge plugin of VMD and with Chimera [172].

#### 2.2.6.3 B-factors

Atomic root mean positional fluctuations (RMSF) were computed with the *ptraj* command *atomicfluct*. The keyword "bfactor" was specified in the script to output data as B-factors ($(8/3)\pi^2$ Å$^2$) instead of RMSF (Å). The scaling factor ($(8/3)\pi^2$ Å$^2$) is "similar but not necessarily equivalent to the calculation of crystallographic B-factors" (AmberTools13, page 239 [7]).

### 2.2.7 Analysis of binding pockets and cavities

A comparison of the structural pockets and cavities was performed by means of CASTp server [173]. CASTp measures areas and volumes of protein pockets and cavities both in solvent accessible surface (SA, Richard's surface) [174] and

[7] http://ambermd.org/doc12/AmberTools13.pdf (June 16, 2013)

molecular surface (MS, Connolly's surface) [175]. In addition, it measures the length of cavities and pockets as well as the number of mouth openings and their area and circumference.

MolAxis [176] was used to identify pathways representing protein molecular channels. Such pathways connect an inner chamber (like a binding site) represented by a single point to the bulk solvent.

The Dowser plugin of VMD [177] was used to locate hydrophilic cavities. The program surveys the structure of a protein to locate internal cavities. It then assesses their hydrophilicity in terms of the energy of interaction of a water molecule with the surrounding atoms.

### 2.2.8 Molecular Interaction Field analysis

The ligand-binding sites of proteins were further characterised by means of an energy-based approach implemented by two software tools, namely, EasyMIFs and SiteHound. These software tools are based on Molecular Interaction Fields (MIFs) that describe the spatial variation of the interaction energy between a probe, usually representing a chemical group and a target protein. For computational convenience, the interaction energy field is discretised around the protein on a three-dimensional orthogonal grid. EasyMIFs provides a MIF calculator the output of which is an energy map of the potential energy between the protein target and a given probe [178]. For the calculations, the program uses the GROMOS [179] force field, *in vacuo*, and a distance dependent dielectric.

SiteHound is a post-processing tool of the EasyMIFs output that enables the prediction of binding site regions that are likely to be involved in the binding of small molecules [180]. The program achieves this by applying an energy filter and,

subsequently, clustering the points of the interaction energy map according to spatial proximity and ranking the clusters by Total Interaction Energy (TIE). The clustering is performed through the application of either single or average linkage agglomerative algorithms.

### 2.2.9 3-D image rendering

All 3-D images were rendered with Chimera [172].

### 2.2.10 Units

Gibbs free energies are reported in kcal/mol. Both AutoDock and Amber report in these units. 1 kcal/mol = 4.2 kJ/mol

*Chapter 3*

# *B. mori & M. sexta OBPs*

## 3.1 Deciphering the specificity of moth OBPs

### 3.1.1 Lepdidopteran sex pheromones and OBPs

Lepidoptera is the second larger insect order and has evolved since the Mesozoic era (>100 million years). It is divided into about 120 families belonging to 46 superfamilies of which 31 consist of moth species. The sex pheromone, which is mostly secreted by the adult female, and much less frequently by the male moth, plays an important role in reproductive isolation and as a result species-specific pheromones differ considerably in their chemical structures. A large number of male and a much smaller number of female attractants corresponding to different moth species have been identified and their structures have been elucidated. In most cases, pheromones secreted by taxonomically related moths are structurally similar. However, structural modifications are necessary to ensure reproductive isolation amongst different species. Additional diversity of the lepidopteran sex pheromones is achieved by the combination of pheromonal components and/or varying the mixing ratio of such combinations. The insect olfactory system shows extreme sensitivity and specificity to different pheromones. It is able not only to detect minute amounts of semiochemicals and, in particular, pheromones, but is also able to discriminate between different stereoisomers and between subtle changes in the composition of odorant blends.

Moth sex pheromones are typically 12-18 carbon partially unsaturated aliphatic hydrocarbons bearing terminal functional groups such as alcohol, aldehyde, or acetate [181]. The structures of the moth pheromones have been classified in detail [182]. About 75% of the known pheromones are secreted by the female moth and are primary alcohols and their derivatives, mainly acetates and aldehydes. They comprise long straight hydrocarbon chains ($C_{10}$-$C_{18}$) of varying

degrees of unsaturation. The pheromones of some of the most notorious pest insects belong to this type of compounds. A second major group of pheromones consist of polyunsaturated hydrocarbons of 17-23 carbon atoms and their epoxy derivatives. This group, comprising some 15% of the known lepidopteran pheromones, lack a functional group at the terminal position of the hydrocarbon chain. A third group of miscellaneous pheromone components consisting of monoenyl and dienyl alcohols of 7-9 carbon atoms, as well as esters between long chain carboxylic acids and shorter chain alcohols have also been isolated and characterised in a number of moth species. Finally, the male moth pheromones have low species-specificity and their structures are very diverse and quite different from those of the female pheromones.

Amongst all insect OBPs, Lepidopteran OBPs show the greatest similarity between species and have been classified into the three classical clusters of OBPs, namely, PBPs, GOBPs and ABPs as discussed in Section 1.3.1. The overall sequence identity and similarity amongst GOBPs is high (31% and 87.5%, respectively). PBPs are much less interrelated (3.9% identity, 69.1% similarity) [93]. Only four structures of Lepidopteran structures have been solved by X-ray crystallography and NMR, at the time of writing of this thesis. Two of the structures solved by solution NMR represent the apoprotein form of *A. polyphemus* and *A. transitella* PBPs, whereas those of *B. mori* PBP1 (*Bmor*PBP1) and *B. mori* GOBP2 (*Bmor*GOBP2) have been co-crystallised with the natural pheromones bombykol and bombykal, as well as with a number of other ligands (Table 1.2, Section 1.4.2).

The crystallographic models of *Bmor*PBP1 have provided evidence that this protein can bind promiscuously diverse hydrophobic molecules in a binding pocket

of considerable plasticity [70]. Binding assays of the pheromone components bombykol and bombykal with *Bmor*PBP1 and *Bmor*GOBP2 have shown that whereas *Bmor*PBP1 was unable to discriminate between the pheromone components, *Bmor*GOBP2, was shown to be more selective for bombykol [64].

In this Chapter, the results of docking and MD simulations are used to gain insights on the ability of *B.mori* and *M. sexta* OBPs to discriminate specific ligands and/or functional groups. A three-stage approach has been followed. In stage one, a relatively large number of compounds of diverse chemical structures were screened for binding affinity to OBPs of *B.mori* and *M. sexta* by means of virtual docking. This was followed by cross docking of bombykol, bombykal as well as a number of their analogues onto these proteins with view to identifying residues that may be involved in ligand recognition. Finally, a more in-depth study of selected OBP/ligand complexes was conducted by means of MD.

# 3.2 Docking simulations

## 3.2.1 Methods

### 3.2.1.1 Pheromones used in virtual docking

Pheromone compounds and their derivatives used in virtual docking are shown in (Supplementary Material: *B. mori* & *M. sexta*: List of compounds contained in the docking library). The compounds were chosen to represent the different types of pheromones described in Section 3.1.1. The criteria used for the selection of these compounds were (a) the presence of polar moieties, (b) a wide range of chain lengths, (c) the position and type of unsaturated bonds. A number of other compounds, mainly benzyl derivatives, were also used. The procedure used for the docking of ligand libraries was that described in Section 2.2.2.2.

### 3.2.1.2 OBPs/PBPs models used in docking simulations

*BmorPBP1* and *Bmor*GOBP2 have been studied in depth and provide the necessary background to validate the results obtained from the methods mentioned above. To test the predictive ability of these methods, two additional structural models were constructed by homology modelling, namely, *M. sexta* PBP1 (*Msex*PBP1) and GOBP2 (*Msex*GOBP2). These two proteins are a reasonable choice as the sequence identity with the corresponding *B. mori* proteins is high despite the fact that the two species are distantly related phylogenetically (Table 3.1). Most interestingly, both species share bombykol and bombykal as pheromone components despite the fact that they share totally different habitats.

| | BmorPBP1 | BmorGOBP2 | MsexPBP1 | MsexGOBP2 |
|---|---|---|---|---|
| BmorPBP1 | 100 | | | |
| BmorGOBP2 | 30.66 | 100 | | |
| MsexPBP1 | 70.80 | 31.21 | 100 | |
| MsexGOBP2 | 30.66 | 79.43 | 30.50 | 100 |

**Table 3.1** **Sequence identity (%) between *B. mori* and *M. sexta* OBPs**

Modeller [183] was used for homology modelling. The sequence UniProt: Q9BMI1 was used to build the model for *Msex*GOBP2 on the template of *Bmor*GOBP2 (PDB: 2wc6). For the *Msex*PBP1 model, UniProt: P18959 was submitted to Modeller using PDB: 1dqe as the template. The two models superimposed on their respective templates are shown in Figures 3.1 and 3.2. The RMSD values and Q-scores against the templates used are given in the Figures. The Q-score values range from zero for completely dissimilar or un-superimposed structures to 1 for identical structures [184].

Figure 3.1 shows that the carboxy-terminal end of the two models is unstructured. Furthermore, the second helix of *Msex*PBP1 is out of register with the one that corresponds to its template (*Bmor*PBP1). Sequence conservation is this region is rather low as shown in the figure. It is noted that residues of this helix do not form part of the binding site of *Bmor*PBP1.

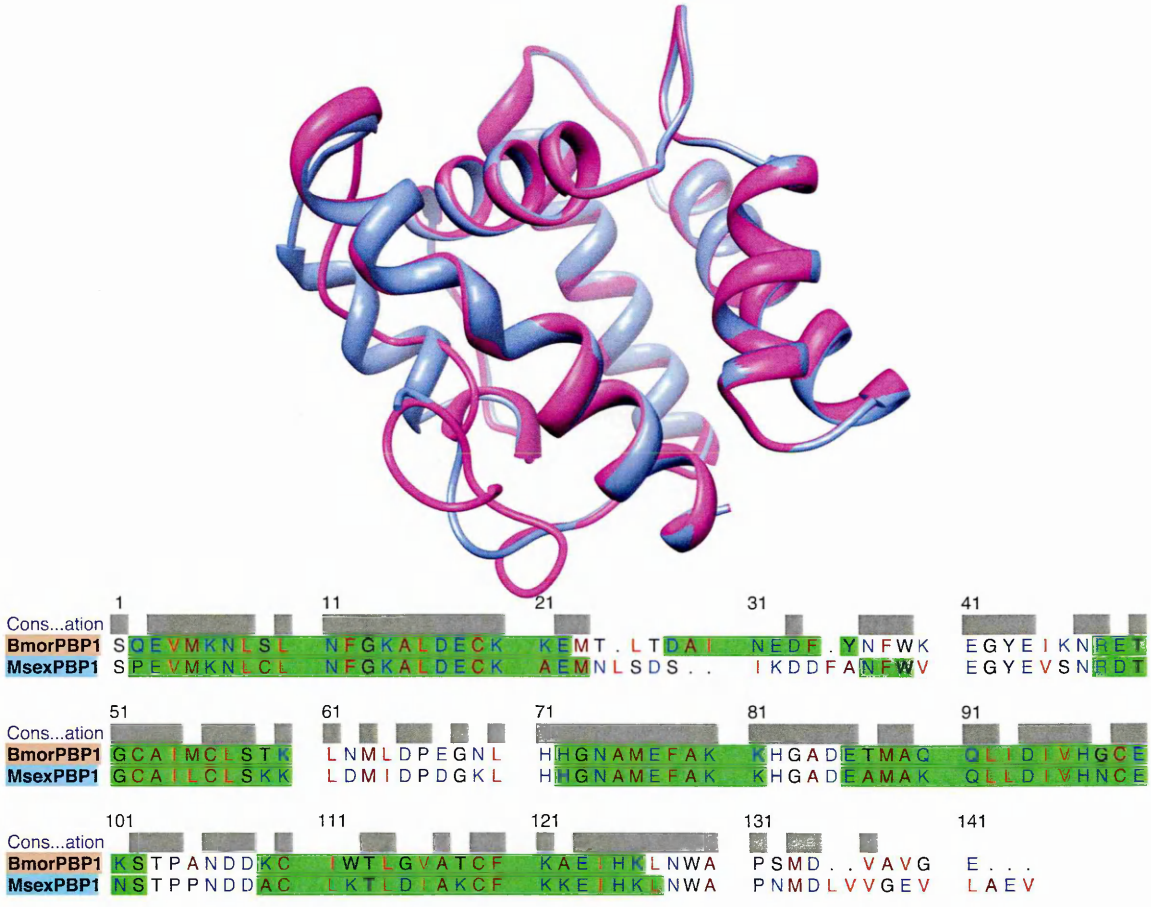



**Figure 3.1** **Structural alignment of *Msex*PBP1 homology model and X-ray model of *Bmor*PBP1**

*Bmor*PBP1: magenta; *Msex*PBP1: blue; Backbone RMSD: 0.692 Å; Q-score: 0.889

The structures of the two models, *Msex*GOBP2 and *Bmor*GOBP2 (Figure 3.2) are nearly identical and the helices are almost perfectly aligned. The carboxy-terminal end of *Msex*GOBP2 is not part of the binding pocket of the protein forming a helix that is directed away from the binding cavity of the protein much like in the X-ray structure of *Bmor*GOBP2.

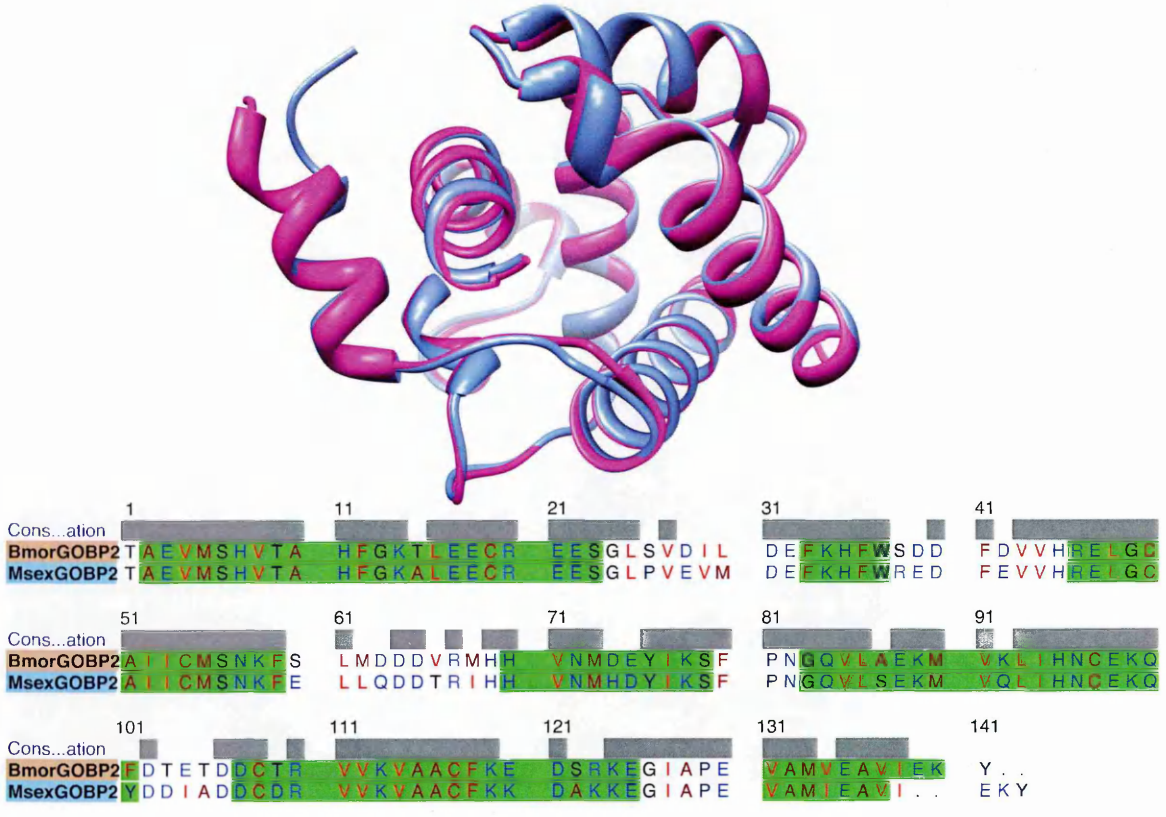



**Figure 3.2** **Structural alignment of *Msex*GOBP2 homology model & X-ray model of *Bmor*GOBP2**

*Bmor*GOBP2: magenta; *Msex*GOBP2: blue; Backbone RMSD: 0.513 Å; Q-score: 0.944

# 3.3 Docking results

## 3.3.1 Re-docking

Re-docking was performed to verify the docking parameters and to recover the structure and interactions of three of the known complexes. These are shown in Figure 3.3 and correspond to *Bmor*PBP1-bombykol (A); *Bmor*GOBP2-bombykol (B); *Bmor*PBP1-bell peper odorant [8] (C), respectively. Bombykol was shown to adopt the same hook-like structure in the two complexes as in the crystallographic models. The redocked ligands maintained the same interactions with residues at the binding sites of the models.

[8] 2-isobutyl-3-methoxypyrazine

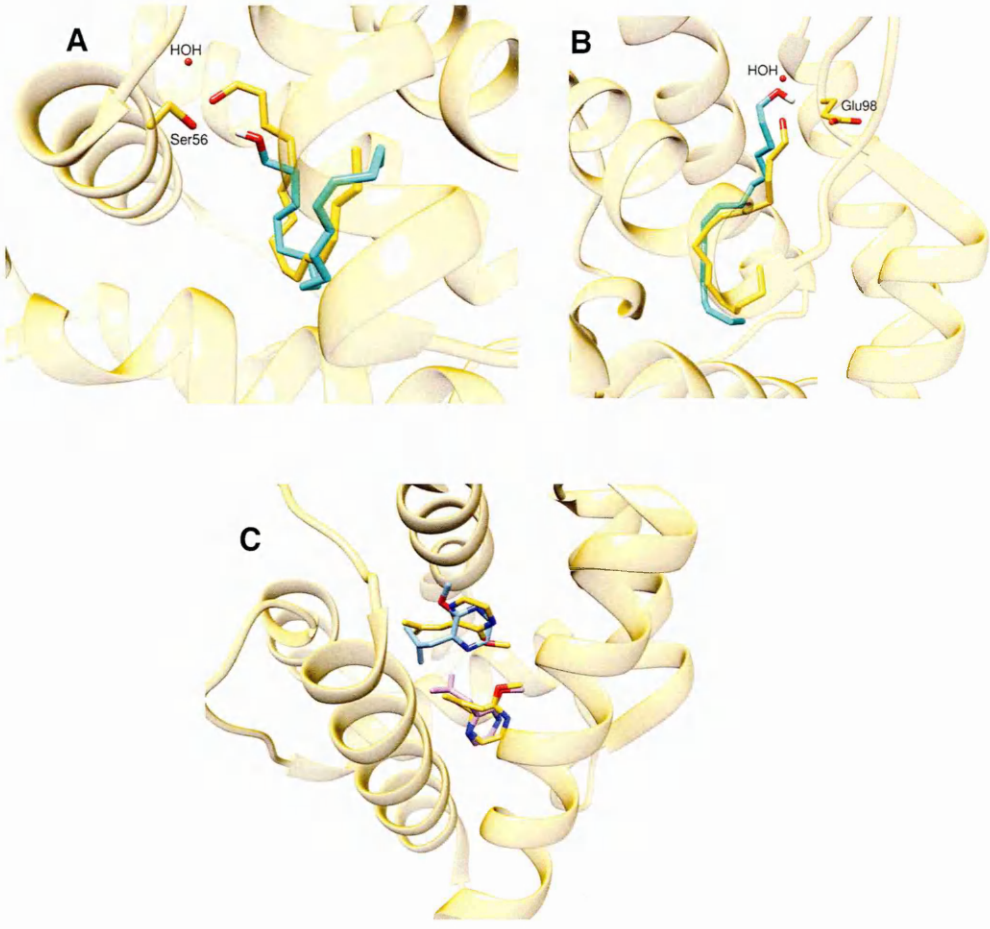



**Figure 3.3** **Re-docking of bombykol and bell pepper odorant on known model structures**

**(A)** 1DQE (*Bmor*PBP1-bombykol); **(B)** X-ray model: 2WC6 (*Bmor*GOBP2-bombykol); **(C)** X-ray model: 2P70 (BmorPBP1-bell pepper odorant).

RMSD of ligands with least squares fitting: 3.9 Å **(A)**; 4.1 Å **(B)**; 2.3 Å and 2.6 Å for each of the two bound ligands **(C)**

### 3.3.2 Screening of ligand libraries

Re-docking is frequently used to verify the suitability of the docking parameters specified in the input files used for virtual screening and cross-docking and to reproduce a known complex's interactions with a particular ligand. Re-docking was performed by taking a number of ligands that have been co-crystalised with OBPs and docking them to the "induced-fit" form of the receptor. These preliminary re-docking simulations established that both the location and poses of the docked ligands were in good agreement with the experimentally solved models. Docking of ligand libraries was used to determine the efficiency of ligand

binding on specific OBPs and thus gain insights as to whether the latter can be used as potential targets for odorant response modulation [66].

The procedure described in this section derives from antecedents in drug discovery and design. Docking scoring functions are well suited for the prediction of protein-ligand binding modes but less well suited for the accurate prediction of binding free energies. In general, for larger hydrophobic ligands, scoring functions have a tendency of underestimating desolvation effects and overestimating van der Waals contacts [185]. It has been found that ligand efficiency is directly related to ligand size. For ligands with up to 15 non-hydrogen atoms, the free energy of binding increases by approximately -1.5 kcal/mol per atom and levels off for ligands with larger relative molecular mass [186]. It is therefore difficult to compare ligand efficiencies for molecules ranging across broad scales of relative molecular mass.

In order to achieve accurate relative ranking of diverse unrelated ligands against a particular receptor ligand efficiency and fit quality indices are used. Ligand efficiency (LE) is any property defining binding affinity, such as $pIC_{50}$, $pK_i$, $\Delta G$, divided by the molecular size of the ligand. Ligand efficiency is thus dependent on molecular size and reaches a plateau for ligands above 45 non-hydrogen atoms. Fit quality (FQ) is an alternative metric that normalises ligand efficiency with respect to size [159].

Virtual screening of ligand libraries was performed with AutoDock Vina as described in Section 2.2.2.2. Ligands and proteins were prepared with the PyRx virtual screening software also described in Section 2.2.2.2. Ligand Efficiency (LE) and Fit Quality (FQ) scores were calculated on the basis of the following formulae derived from Bembenek *et al.* [159]:

$$LE = \Delta G / HA \qquad (HA = \text{heavy atom count}) \tag{3.1}$$

$$FQ = LE / LE\_Scale \tag{3.2}$$

$$LE\_Scale = 0.0715 + \frac{7.5328}{HA} + \frac{25.7079}{HA^2} + \frac{361.4722}{HA^3} \tag{3.3}$$

For a given atom number, the best ligand efficiencies are those whose FQ tends to 1.0 whereas poor ligand efficiencies are those with FQ indices less than 0.5.

The free energies of binding of the ten ligands showing the highest affinity for the *B. mori* and *M. sexta* OBPs are shown in Table 3.2. Ligand IDs as well as the detailed results of virtual screening are included in (Supplementary Material: *B. mori* & *M. sexta*: Results of virtual screening).

| ***Bmor*PBP1** | | ***Bmor*GOBP2** | | ***Msex*PBP1** | | ***Msex*GOBP2** | |
|---|---|---|---|---|---|---|---|
| **Ligand ID** | $\Delta G$ (kcal mol$^{-1}$) | **Ligand ID** | $\Delta G$ (kcal mol$^{-1}$) | **Ligand ID** | $\Delta G$ (kcal mol$^{-1}$) | **Ligand ID** | $\Delta G$ (kcal mol$^{-1}$) |
| 92 | -10 | 93 | -9.4 | 82 | -6.7 | 93 | -9.6 |
| 93 | -9.4 | 92 | -9.2 | 91 | -6.7 | 92 | -8.8 |
| 91 | -8.3 | 91 | -8.1 | 81 | -6.6 | 63 | -7.7 |
| 82 | -8.2 | 82 | -7.8 | 90 | -6.5 | 65 | -7.5 |
| 81 | -7.8 | 81 | -7.7 | 88 | -6.3 | 81 | -7.4 |
| 90 | -7.8 | 63 | -7.7 | 80 | -6.3 | 82 | -7.4 |
| 65 | -7.8 | 90 | -7.6 | 79 | -6.2 | 66 | -7.4 |
| 64 | -7.8 | 19 | -7.5 | 70 | -6.2 | 19 | -7.4 |
| 27 | -7.8 | 28 | -7.5 | 24 | -6.2 | 20 | -7.3 |
| 66 | -7.6 | 27 | -7.5 | 89 | -6.2 | 91 | -7.2 |

**Table 3.2** **Free energies of binding of ligands in the screening library**

The data shown in the table indicate that, in general, the free energies of binding of the complexes increase in the order:

*Bmor*PBP1 ≈ *Bmor*GOBP2 ≥ *Msex*GOBP2 > *Msex*PBP1

Of the ten ligands showing the highest affinities, two are common to all four proteins and five common to *Bmor*PBP1, *Bmor*GOBP2 and *Msex*GOBP2. The

highest binding affinities are reported for polyunsaturated alcohols linked to an aromatic or cyclohexane ring, bicyclic compounds such as caryophyllene and unsaturated epoxy compounds. Figure 3.4 shows the four ligands of the highest free energies of binding for each protein.

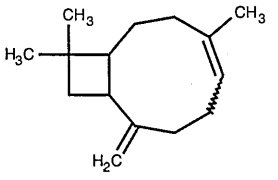
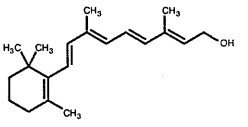
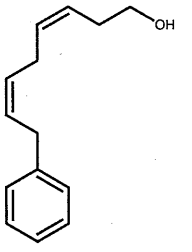
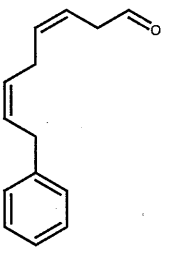
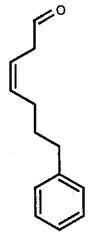
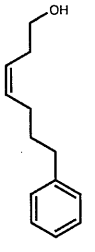
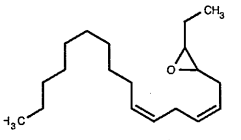
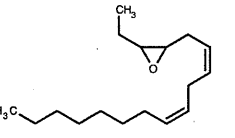
| | | | |
|---|---|---|---|
|  |  |  |  |
| **Lig1.** 4,11,11-trimethyl-8-methylidenebicyclo[7.2.0]undec-4-ene | **Lig2.** (2E,4E,6E, 8E)-3,7-dimethyl-9-(2,6,6-trimethylcyclohex-1-en-1-yl)nona-2,4,6,8-tetraen-1-ol | **Lig3.** (3Z,6Z)-8-phenylocta-3,6-dien-1-ol | **Lig4.** (3Z,6Z)-8-phenylocta-3,6-dienal |
| BmorPBP1: -10.0 (1)<br>BmorGOBP2: -9.4 (2)<br>MsexPBP1: -5.4 (76)<br>MsexGOBP2: -8.8 (2) | BmorPBP1: -9.4 (2)<br>BmorGOBP2: -9.2 (1)<br>MsexPBP1: -6.2 (7)<br>MsexGOBP2: -9.6 (1) | BmorPBP1: -8.3 (3)<br>BmorGOBP2: -8.1 (3)<br>MsexPBP1: -6.7 (1)<br>MsexGOBP2: -7.2 (10) | BmorPBP1: -8.2 (4)<br>BmorGOBP2: -7.8 (4)<br>MsexPBP1: -6.7 (2)<br>MsexGOBP2: -7.4 (6) |
|  |  |  |  |
| **Lig5.** (3Z)-7-phenylhept-3-enal | **Lig6.** (3Z)-7-phenylhept-3-en-1-ol | **Lig7.** 2-ethyl-3-[(2Z,5Z)-pentadeca-2,5-dien-1-yl]oxirane | **Lig8.** 2-ethyl-3-[(2Z,5Z)-trideca-2,5-dien-1-yl]oxirane |
| BmorPBP1: -7.8 (5)<br>BmorGOBP2: -7.7 (5)<br>MsexPBP1: -6.6 (3)<br>MsexGOBP2: -7.4 (6) | BmorPBP1: -7.8 (6)<br>BmorGOBP2: -7.6 (7)<br>MsexPBP1: -6.5 (4)<br>MsexGOBP2: -7.1 (12) | BmorPBP1: -7.4 (16)<br>BmorGOBP2: -7.7 (6)<br>MsexPBP1: -5.4 (77)<br>MsexGOBP2: -7.7 (3) | BmorPBP1: -7.8 (5)<br>BmorGOBP2: -7.4 (11)<br>MsexPBP1: -5.3 (84)<br>MsexGOBP2: -7.7 (4) |

**Figure 3.4** **Ligands showing the highest free energies of binding**

Free energies of binding are in kcal $mol^{-1}$. Numbers in parentheses represent the relative ranking of the ligands

Of the ligands shown in the above figure, the aromatic compounds and retinol (Lig2) do not belong to the odour space of Lepidoptera, caryophyllene (Lig1) is known to elicit olfactory response in moths and the monoepoxy diene compound is

an analogue of the female pheromones of Noctuidae and Lymantriidae, but not of the Bombycidae and Sphengidae species to which the tested proteins belong.

*Bmor*PBP1 scored highest binding energy with caryophyllene, *Bmor*GOBP2 and *Msex*GOBP2 with retinol and *Msex*PBP1 with (*3Z,6Z*)-8-phenylocta-3,6-dien-1-ol. Of the four OBPs tested, *Msex*PBP1 showed the weakest binding.

Free energies of binding were converted into fit quality scores in order to obtain a better picture of binding affinities of the docked ligands against their respective receptors. The FQ scores of the ligands docked on each protein are shown in Figure 3.5. The trend lines representing the linear regression between the FQ scores and the number of ligand heavy atoms (HA) show a tendency of decreasing FQ score with HA. However, significant correlation of FQ with the heavy atom count (HA) is shown only in the case of *Msex*PBP1. With the exception of *Msex*GOBP2, all other proteins seem to favour ligands with 6-10 heavy atoms.

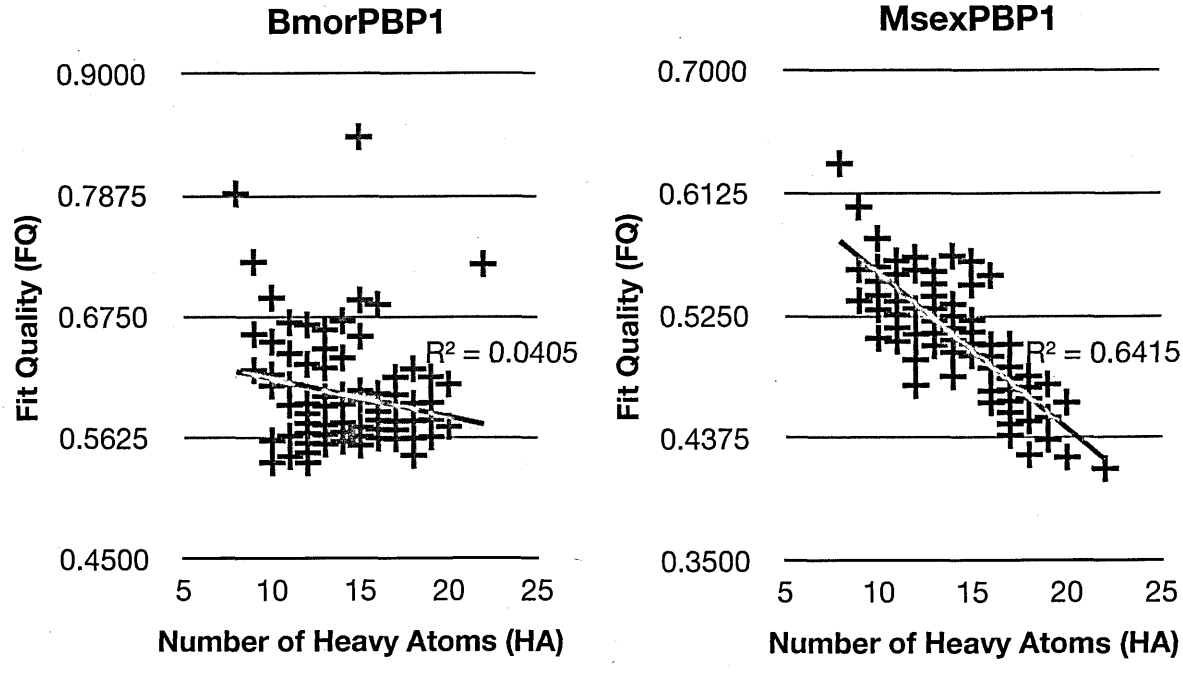



**Figure 3.5** **Ligand FQ scores for the *B. mori* and *M. sexta* OBPs (contd.)**

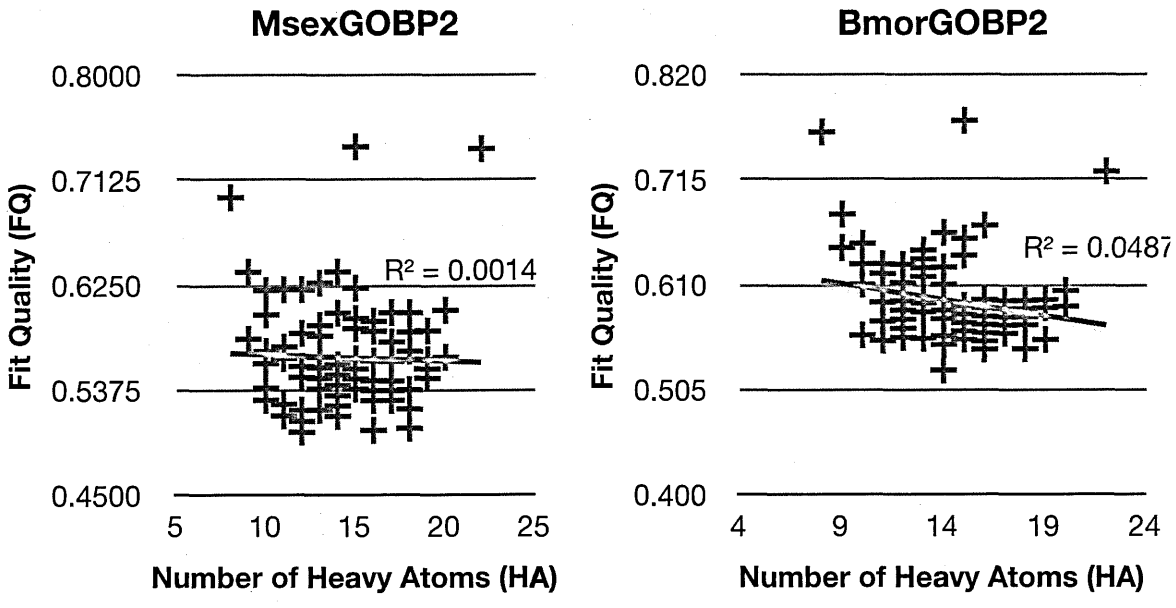



**Figure 3.5 (contd.)** **Ligand FQ scores for the *B. mori* and *M. sexta* OBPs**

Figure 3.6 shows the four highest FQ-scoring ligands for each protein.

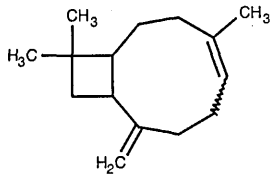
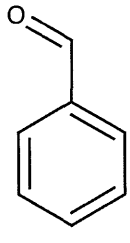
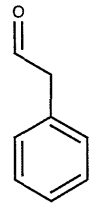
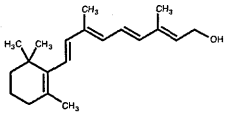
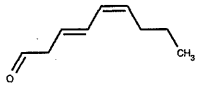
| | | |
|---|---|---|
|  |  |  |
| **Lig1**. Caryophyllene* | **Lig2.** Benzaldehyde | **Lig3.** 2-phenylacetaldehyde |
| BmorPBP1 FQ = 0.84 rank: 1<br>BmorGOBP2 FQ = 0.77 rank: 1<br>MsexPBP1 FQ = 0.52 rank: 2<br>MsexGOBP2 FQ = 0.74 rank: 1 | BmorPBP1 FQ = 0.79 rank: 2<br>BmorGOBP2 FQ = 0.76 rank: 2<br>MsexPBP1 FQ = 0.63 rank: 1<br>MsexGOBP2 FQ = 0.70 rank: 3 | BmorPBP1 FQ = 0.73 rank: 3<br>BmorGOBP2 FQ = 0.68 rank: 4<br>MsexPBP1 FQ = 0.60 rank: 2<br>MsexGOBP2 FQ = 0.64 rank: 5 |
|  |  |  |
| **Lig4.** Retinol* | **Lig5.** (3E,5Z)-nona-3,5-dienal | **Lig6.** (3Z)-7-phenylhept-3-en-1-al |
| BmorPBP1 FQ = 0.72 rank: 4<br>BmorGOBP2 FQ = 0.72 rank: 3<br>MsexPBP1 FQ = 0.32 rank: 93<br>MsexGOBP2 FQ = 0.74 rank: 2 | BmorPBP1 FQ = 0.61 rank: 27<br>BmorGOBP2 FQ = 0.61 rank: 22<br>MsexPBP1 FQ = 0.58 rank: 3<br>MsexGOBP2 FQ = 0.57 rank: 33 | BmorPBP1 FQ = 0.67 rank: 8<br>BmorGOBP2 FQ = 0.66 rank: 6<br>MsexPBP1 FQ = 0.57 rank: 4<br>MsexGOBP2 FQ = 0.63 rank: 4 |

**Figure 3.6** **Highest FQ-scoring ligands in the screening library**

* IUPAC name is given in Figure 3.5

The common features of these compounds are (a) the presence of aromatic and/ or cyclic rings and (b) the presence of conjugated double bonds. These features are shared by most of the highest FQ-scoring ligands (Supplementary Material: *B. mori* & *M. sexta*: Results of virtual screening). It is noted that the highest FQ-score for *Msex*PBP1 are given by aldehydes and are considerably lower than the highest FQ-scores for the other proteins.

These results should be treated with caution as LE and Fit Quality scores are empirical and have been parameterised mainly on different types of pharmacophores rather than on the type of compounds representative of the Lepidopteran odour space. Nevertheless, some coarse-grained conclusions can be drawn. Whereas there is little to differentiate *Bmor*PBP1, *Bmor*GOBP2 and *Msex*GOBP2 in terms of binding affinities for the type of ligands present in the docking library, *Msex*PBP1 appears to favour ligands with different structural characteristics. It is not possible to conclude whether this apparent differentiation of *Msex*PBP1 from the other proteins is "real" or due to structural deficiencies of the constructed homology model.

### 3.3.2 Cross docking

Cross-docking was performed to discern finer details with regard to ligand selectivity by the OBPs. The crystallographic models of the *BmorPBP1*, 1DQE [58] and of the *Bmor*GOBP2, 2WC6 [64] were used after stripping them from water and bound ligands. The corresponding *M. sexta* proteins were solvated, energy minimised and equilibrated for 2ns as described in Section 2.2.3.1. The minimum energy structures were extracted from the 2ns equilibration trajectory, stripped of water molecules and used for docking.

The ligands used in cross docking simulations are shown in Figure 3.7. They include the highest FQ-scoring ligands as well as the pheromone components bombykol and bombykal. Docking simulations were performed as described in Section 2.2.2.1. The resulting families of docked conformations were clustered with a 2 Å RMSD tolerance, and were ranked in order of increasing energy.

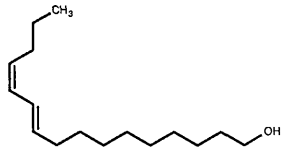
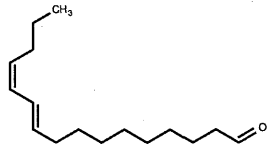
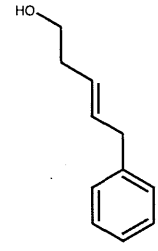
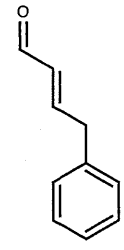
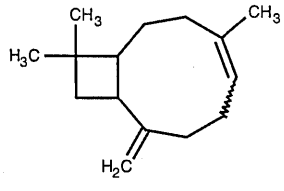
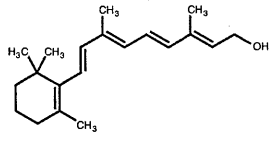
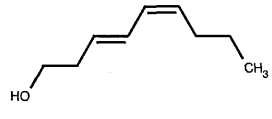
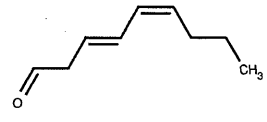
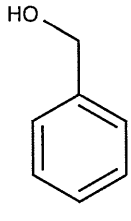
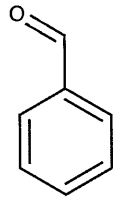
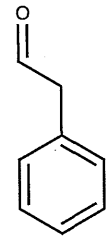
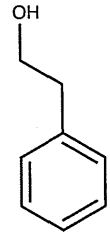
| | | | |
|---|---|---|---|
|  **Lig1:** (10E,12Z)-hexadeca-10,12-dien-1-ol |  **Lig2:** (10E,12Z)-hexadeca-10,12-dienal |  **Lig3:** (3E)-5-phenylpent-3-en-1-ol |  **Lig4:** (3E)-5-phenylpent-3-en-1-al |
|  **Lig5:** 4,11,11-trimethyl-8-methylidenebicyclo[7.2.0]undec-4-ene (β-caryophyllene) |  **Lig6:** (2E,4E,6E,8E)-3,7-dimethyl-9-(2,6,6-trimethylcyclohex-1-en-1-yl)nona-2,4,6,8-tetraen-1-ol (retinol) |  **Lig7:** (3E,5Z)-nona-3,5-dien-1-ol |  **Lig8:** (3E,5Z)-nona-3,5-dienal |
|  **Lig9:** phenylmethanol |  **Lig10:** benzaldehyde |  **Lig11:** 2-phenylacetaldehyde |  **Lig12:** 2-phenylethanol |

**Figure 3.7** **Ligands used in cross docking simulations**

Table 3.3 shows the energies and residues forming H-bonds of the minimum binding energy ligand clusters.

| | *BmorPBP1* (kcal/mol) | | | | *MsexPBP1* (kcal/mol) | | | |
|---|---|---|---|---|---|---|---|---|
| | **No. Clust.** | **Freq. (%)** | **$E_{min}$** | **H-bonds** | **No. Clust.** | **Freq. (%)** | **$E_{min}$** | **H-bonds** |
| **Lig1** | 11 | 37 | -7.16 | - | 13 | 31 | -6.4 | T111 |
| **Lig2** | 6 | 29 | -7.6 | - | 5 | 43 | -6.5 | - |
| **Lig3** | 3 | 80 | -5.7 | S9; W37 | 2 | 82 | -5.3 | - |
| **Lig4** | 3 | 65 | -5.7 | - | 3 | 95 | -5.6 | - |
| **Lig5** | 1 | 100 | -8.4 | - | 2 | 99 | -7.1 | - |
| **Lig6** | 3 | 30 | -10.2 | E98 | 15 | 32 | -5.7 | - |
| **Lig7** | 4 | 83 | -4.8 | T9 | 3 | 49 | -4.6 | S56 |
| **Lig8** | 4 | 52 | -4.9 | - | 2 | 44 | -4.8 | - |
| **Lig9** | 1 | 100 | -4.3 | - | 2 | 22 | -4.4 | K6 |
| **Lig10** | 1 | 100 | -4.3 | - | 2 | 24 | -4.3 | - |
| **Lig11** | 1 | 100 | -4.6 | - | 2 | 90 | -4.5 | - |
| **Lig12** | 3 | 97 | -4.6 | - | 4 | 2 | -4.5 | - |

| | *BmorGOBP2* (kcal/mol) | | | | *MsexGOBP2* (kcal/mol) | | | |
|---|---|---|---|---|---|---|---|---|
| | **No. Clust.** | **Freq. (%)** | **$E_{min}$** | **H-bonds** | **No. Clust.** | **Freq. (%)** | **$E_{min}$** | **H-bonds** |
| **Lig1** | 13 | 40 | -7.6 | E98 | 16 | 14 | -7.1 | E98 |
| **Lig2** | 7 | 44 | -7.7 | R110 | 7 | 33 | 7.0 | - |
| **Lig3** | 3 | 48 | -5.7 | - | 11 | 11 | -5.4 | S56 |
| **Lig4** | 2 | 96 | -5.8 | - | 7 | 11 | -5.4 | S56 |
| **Lig5** | 1 | 100 | -8.6 | | 1 | 100 | -8.2 | - |
| **Lig6** | 1 | 100 | -10.2 | E98 | 2 | 86 | -8.85 | R110 |
| **Lig7** | 5 | 83 | -4.8 | T9 | 7 | 49 | -4.8 | I68 |
| **Lig8** | 3 | 89 | -5.0 | - | 6 | 67 | -4.7 | - |
| **Lig9** | 2 | 91 | -4.2 | T9 | 2 | 93 | -4.2 | E32 |
| **Lig10** | 1 | 100 | -4.3 | - | 1 | 100 | -4.2 | S56 |
| **Lig11** | 2 | 99 | -4.6 | - | 1 | 100 | -4.5 | S56 |
| **Lig12** | 5 | 9 | -4.3 | - | 3 | 96 | -4.4 | R110 |

**Table 3.3 H-bond formation and binding energies of the lowest energy clusters**

No. Clust. is the number of distinct multi-member conformational clusters resulting from 100 runs; Freq. (%) is the per cent population of the lowest energy cluster; $E_{min}$ is the free energy of binding of the most favourable conformation of the lowest energy cluster

The binding energies were dominated predominantly by hydrophobic and van der Waals interactions giving rise to favourable desolvation energies. The contribution of electrostatic interactions to the binding energies was, in all cases, less than

1 kcal $mol^{-1}$. No definitive conclusion can be drawn with regard to the relative binding affinities of the proteins to the smaller aromatic ligands as the reported free energies of binding are well within the standard deviation of the AutoDock force field (2.5 kcal $mol^{-1}$). However, a certain trend of binding affinities is observed in which the free energy of binding of the *B. mori* OBPs-alcohol complexes is lower than that of the *M. sexta* OBPs.

*Msex*GOBP2 is, likewise, shown to bind alcohols with greater affinity than *Msex*PBP1 but less than *Bmor*PBP1 and *Bmor*GOBP2. With the exception of bombykal, the situation is reversed for the aldehyde ligands. *Msex*PBP1 is shown to bind aldehyde ligands with almost equal or better affinity to the *B. mori* OBPs and with considerably more affinity than *Msex*GOBP2. This is consistent with observation obtained from the docking of the ligand library mentioned earlier.

The results obtained from docking caryophyllene and retinol stand out and, in the case of the latter, the binding energy lies outside the error margins of the AutoDock force field. The free energies of binding between the two compounds and the *B. mori* proteins was by far lower than any ligand tested. Both molecules are hydrophobic, retinol less so, but their geometries are very different. Unlike caryophyllene which has a compact and rigid structure, retinol possesses a cyclic structure in the form of a $\beta$-ionone ring which is attached to a stretched polyene chain with four conjugated double bonds. Caryophyllene docked on each of the proteins in a single cluster. Binding is mediated totally by hydrophobic interactions as the molecule has no polar moieties attached to it. The table shows that binding to *Msex*PBP1 is much weaker than the other proteins. Binding to *Msex*GOBP2 is also much weaker than to the *B. mori* OBPs. The free energies of binding of the retinol complexes are strikingly different between the *B. mori* and the *M. sexta*

proteins. Binding to the latter is considerably weaker and in the case of MsexPBP1, uncharacteristically weak.

H-bonds reported in Table 3.3 refer to the lowest energy ligand clusters. In the crystallographic models of the *Bmor*PBP1- and *Bmor*GOBP2-bombykol complexes, H-bond formation between the bombykol and the proteins involves hydroxyl group of bombykol and Ser56 of *Bmor*PBP1, and Glu98 and Arg110 *Bmor*GOBP2. In both complexes, water molecules are involved in bridged H-bonds. In earlier docking work with *Bmor*PBP1 and *Bmor*GOBP2 bombykol analogues in which the conjugated (10$E$, 12$Z$) double bonds were moved to positions (8$E$, 10$Z$), the free energy of binding between the ligands and *Bmor*PBP1 and *Bmor*GOBP2 was reduced. Similar reduction in binding energy was observed on docking analogues of bombykol with unsaturated triple bonds. Furthermore the lowest energy minima reached for the GOBP-ligand complexes were consistently lower than the corresponding PBP ones and poor clustering of ligand conformations was also observed [187].

Table 3.3 also shows that for the aromatic or cyclic alcohols with much fewer rotatable bonds than bombykol clustering was much better and converged to the lowest energy solution. H-bonds between the ligands and the proteins, in the lowest energy clusters, involved mainly Ser9 for *Bmor*PBP1; Lys6 and in one case Ser56 for *Msex*PBP1; Thr9, Arg67, Glu98 and R110 for *Bmor*GOBP2 and Ser56, Arg67, Glu98 and Arg110 from *Msex*GOBP2. The corresponding less polar aldehydes formed H-bonds with the same residues or none at all.

The clustering of the docked conformations of both bombykol and bombykal was found to be rather poor. This is hardly surprising as these ligands have 12 rotatable bonds with a concomitantly large number of torsional degrees of

freedom. This represents a major algorithmic challenge for docking programs [188]. The poor clustering of the docked structures is in agreement with the findings of Gräter *et al.* who conducted docking simulations of bombykol with *BmorPBP1* and likewise observed poor clustering [98]. In addition, the authors observed many-fold pheromone conformations forming different hydrogen bonds with the protein, suggestive of a highly flexible ligand within the binding pocket. This is consistent with the findings of this work in which members of different conformation clusters of bombykol and bombykal showed H-bond formation with different residues in the binding sites of all four protein receptors. An example of this is illustrated in Figure 3.8, which shows the docking histogram of bombykol conformation clusters in the binding cavity of *Bmor*PBP1, as well as the observed H-bonds for the major clusters.

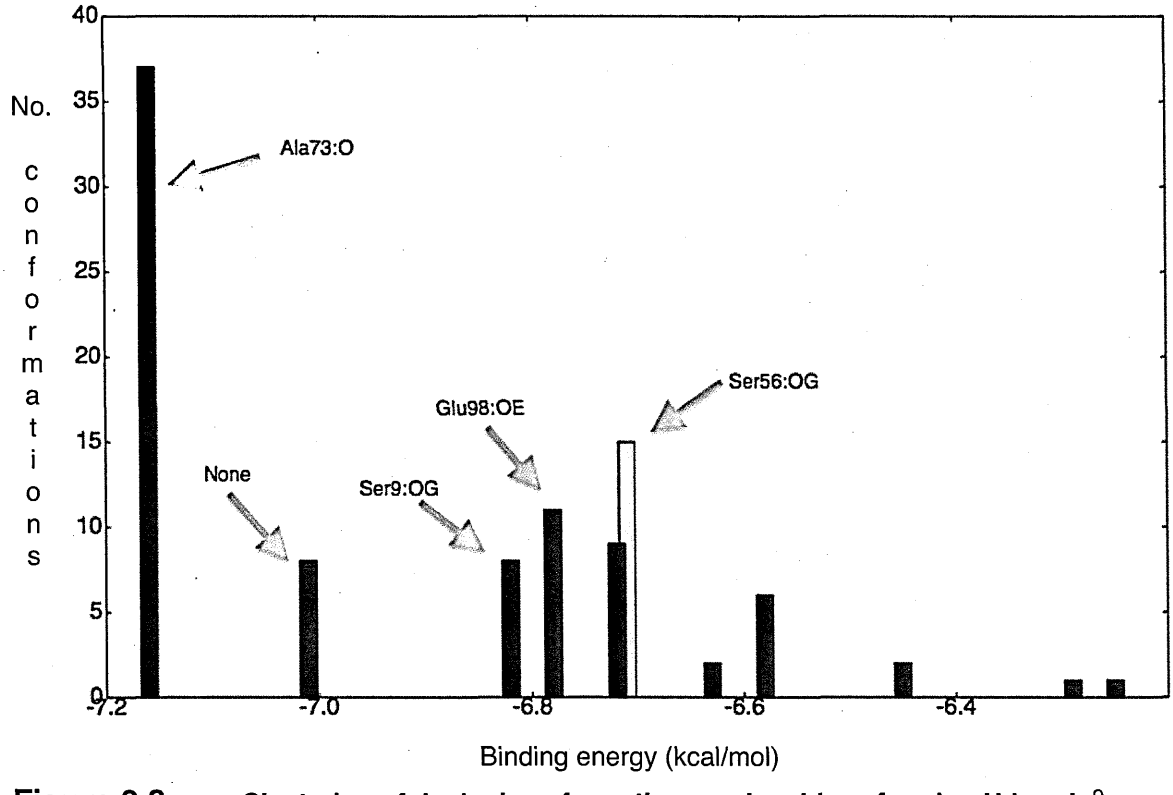



**Figure 3.8** **Clustering of docked conformations and residues forming H-bonds**[9]

The conformations of bombykol in the cluster depicted in red are closest to the X-ray model. Conformation clusters are ranked in order of increasing energy (rank 1 is the lowest energy cluster

[9] Not all members of a given cluster form H-bonds

It can be seen in the figure that the docking cluster in which bombykol forms H-bonds with Ser56 and has conformations that are closest to the the X-ray model is neither the most populated nor of the lowest energy. Similar observations were made with docking simulations involving bombykol analogues in which the hydrocarbon double bonds were in different positions and/or the chain length varied (results not reported). It was also observed that doubling the number of energy evaluations per docking run (200 instead of 100) did not result in significant improvement of clustering. For ligands with large number of torsions, such as most of the ones used in this work, the dimensionality of the search problem is such that a far larger number of evaluations would be required. To some extent the limitations arising from poor clustering and the choice of the energy clusters representing ligand conformations of biological relevance can be overcome by visual inspection and chemical judgement.

The results summarised in Table 3.3 suggest that there are two distinct sites in the binding pocket of the moth OBPs with high affinity for polar groups. One of these is close to the residues seen to form H-bonds in the X-ray models (i.e. Ser56, Glu98 and Arg110), and the other close to the residues in positions 9 and 37. The relative position of these residues with respect to each other are shown in Figure 3.9. The figure shows bombykol in two distinct conformations obtained from the energy clusters in which H-bonding with Ser9 and Ser56 was shown to occur (clusters ranked 3 and 6, respectively, in Figure 3.8 above). It is therefore possible that these OBPs can bind simultaneously more than one ligand. The crystallographic model of *Bmor*PBP1 in complex with the bell pepper odorant (see Figure 3.3 C) supports this proposition [70]. In this model, two bell pepper odorant molecules were shown in the binding site of *Bmor*PBP1. In one of these the methoxy-oxygen

of the molecule lies 3.8 Å from the side chain oxygen of Ser56, whereas the second molecule of the odorant within 4.0 Å of residues Ser9, Phe12 and Phe36.

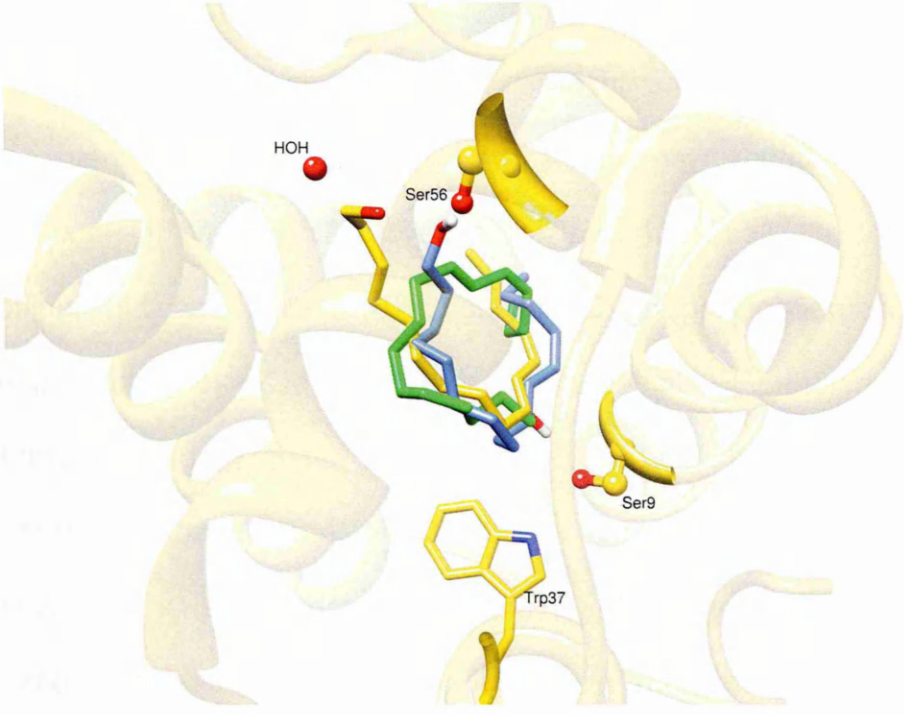



**Figure 3.9** **Distinct conformations of bombykol in the binding site of *Bmor*PBP1**
gold: X-ray model; blue: docking cluster 6 (Figure 3.8); green: docking cluster 3 (Figure 3.8)

The docking results presented in this thesis and the X-ray *Bmor*PBP1-bell pepper model suggest the presence of two distinct low energy "hot spots" within the binding site of the studied proteins, one proximal to residues 56, 98, 110 and the other proximal to residues 37 and 9. Binding of ligands containing polar moieties is equally favourable energetically at these "hot spots".

The capacity of these proteins to bind simultaneously two ligands was tested by docking bell pepper odorant on the BmorPBP1 in which one of the two "hot spots" was occupied by a ligand molecule. The binding energies were measured and were compared against the corresponding ones derived from the docking of the same ligand on the respective apoprotein. The results are summarised in Table 3.4.

| Ligand | X-ray model | Receptor | $\Delta G$ (kcalmol$^{-1}$) |
|---|---|---|---|
| bell pepper | 2fjy | apo*Bmor*PBP1 | -4.6 |
| bell pepper | 2p70 | BP1*Bmor*PBP1 | -5.0 |
| bell pepper | 2p70 | BP2*Bmor*PBP1 | -4.9 |
| bell pepper | 1dqe | bom*Bmor*PBP1 | -3.0 |

**Table 3.4 Re-docking of bell pepper odorant onto *Bmor*PBP1**

bell pepper: 2-isobutyl-3-methoxypyrazine; apo*Bmor*PBP1: *Bmor*PBP1 apoprotein; BP1*Bmor*PBP1: *Bmor*PBP1- bell pepper complex in which the ligand is proximal to Ser56; BP2*Bmor*PBP1: similar to BP1*Bmor*PBP1 in which the ligand is proximal to S9; bom*Bmor*PBP1: *Bmor*PBP1-bombykol complex

The free energies of binding of bell pepper and the BP1(BP2)*Bmor*PBP1 complexes were almost identical. They were ~0.4 kcal/mol lower than the free energy of binding of bell pepper - apo*Bmor*PBP1 complex. Although this could possibly suggest co-operative binding, the result is inconclusive as the difference in free energies of binding was within the standard deviation of the AutoDock force field. Co-operative binding was also excluded in the X-ray study [70] .Docking of bell pepper on the *Bmor*PBP1-bombykol complex resulted in a much higher energy of binding, indicative of steric clashes between the bell pepper and bombykol molecules. This would suggest that whereas *Bmor*PBP1 can accommodate more than one small molecule within its binding cavity, it would not tolerate a second ligand molecule when in complex with its physiological pheromone. A possible implication of this is that given the high concentration of PBPs in the sensillum lymph and the fact that only fraction of these proteins is needed to transport the pheromone to the sensory neurons, the vast majority of the PBP molecules may act to sequester other odorant molecules. This would support the "transporter-scavenger" model of OBP function.

On the basis of what has described above, it seems that the relative orientation (pose) of a given ligand in the binding site of these proteins may not only be

determined by the presence of polar groups alone. Given that H-bond formation makes a minor contribution to the free energy of binding which is dominated by hydrophobic interactions, the presence of unsaturated bonds and/or aromatic moieties may be most important in stabilising the ligands in a particular pose though cation-π and π-π interactions with the aromatic side chains of the conserved Phe12, Phe33 and Phe36 that line the binding sites of the proteins. Notwithstanding the limitations of docking simulations, as well as the possible effect of buried water molecules on binding selectivity that have not been accounted for, all other evidence tends to support the notion that these proteins are able to accommodate more than one small ligand simultaneously.

### 3.3.3 Energetic determinants of binding specificity

To gain further insights on the observed hydrogen-bonding patterns described above, the ligand-binding sites of the four proteins were further characterised by means of an energy-based approach implemented by two software tools, namely, EasyMIFs and SiteHound (see Section 2.2.8). Hydroxyl (OA), phosphate oxygen (OP) and methyl-carbon (CMET) probes were used to gain insights regarding high binding affinity sites within the binding pockets of the proteins. The Average Linkage algorithm was chosen as it produced more discrete energy clusters than the Single Linkage algorithm. The use of the OP probe produced TIE clusters adjacent to or overlapping with those obtained from the use of the OA probe but of higher energy an smaller volume. The results below summarise the findings resulting from the use of the OA and the CMET probes.

#### 3.3.3.1 *B. mori* and *M. sexta* PBPs

The two highest ranking clusters of the OA probe were located in two distinct regions of the proteins as shown in Table 3.5. The residues within 4 Å of each cluster are also shown.

| | Cluster rank | TIE (kcal mol-1) | Volume (Å3) | Cluster surrounding residues |
|---|---|---|---|---|
| *Bmor*PBP1 | 1 | -219 | 68 | F12; F36; I52; S56; M53; M61; L62; G66; L68; V94; E98; W110; T111; V114; A115; F118 |
| | 2 | -184 | 57 | M5; L8; S9; F12; F33; F36; W37; I52; F76; A115; F118; V135 |
| *Msex*PBP1 | 1 | -352 | 124 | M5; L8; F12; F33; W37; I52; L55; S56; M61; A73; A77; H80; L90; V94; E98; K110; I114; A115; F118 |

**Table 3.5** **Moth PBPs: regions of high TIE between the OA probe and the proteins**

A comparison of the two PBPs shows that in *Msex*PBP1 there is a single region of very high total interaction energy (TIE). This region encompasses Ser56 which, as discussed earlier, were shown to form H-bonds with mainly hydroxylated ligands. In *Bmor*PBP1, there are two distinct regions in which the OA probes cluster with comparable TIE, albeit considerably lower than that of the highest energy cluster of *Msex*PBP1. The residues surrounding these cluster regions include Ser9, and the conserved Phe12, Phe33, Phe36 and Trp37 which have also been shown to form H-bonds with docked ligands. Although the volumes of the regions of high TIE are comparable in the two proteins they do not overlap entirely and there are specific regions of TIE in which interactions are more favourable for one protein relative to the second one. These are depicted in Figure 3.10 where it can be seen that the two OA energy clusters in the binding cavity of *Bmor*PBP1 are contiguous. The highest energy cluster is close to Ser56 and Glu98 whereas the lower energy cluster lies close to Ser9 and Trp37. These two clusters overlap to some extent with the highest energy cluster of *MexPBP*1 part of which is also surrounded by Ser56, and Trp37 and Glu98. In *Msex*PBP1, Ser9 is replaced by Cys9.

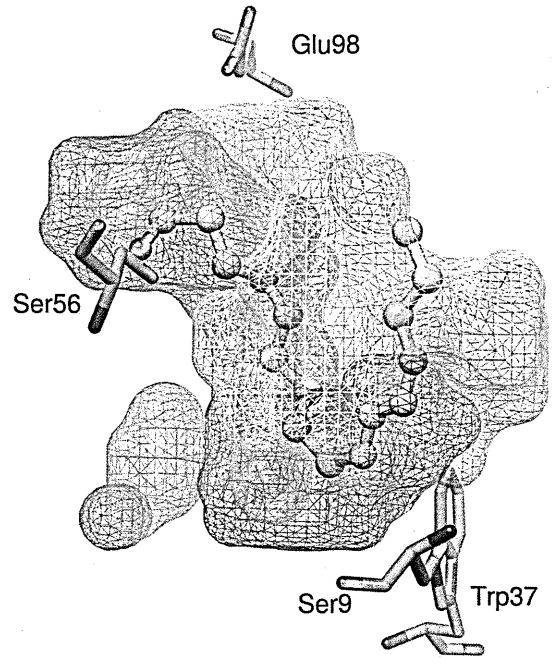



**Figure 3.10** **Moth PBPs: TIE clusters between the proteins and the OA probe**

The X-ray protein models have been superimposed. The *Bmor*PBP1 energy clusters are depicted red (highest TIE) and blue (lower TIE). The *Msex*PBP1 highest energy cluster is shown in yellow. A low energy outlier is shown in green

The results of probing the binding sites with the methyl-carbon probe (CMET) are summarised in Table 3.6.

| | Cluster rank | TIE (kcal mol-1) | Volume (Å3) | Cluster surrounding residues |
|---|---|---|---|---|
| *Bmor*PBP1 | 1 | -213 | 56 | M5; L8; S9; F12; F33; F36; W37; A73; F76; A77; L90; I91; V94; F118; V135 |
| | 2 | -200 | 55 | F12; F36; I52; M53; S56; M61; L62; G66; N67; L68; V94; E98; W110; T111; V114; A115; F118 |
| *Msex*PBP1 | 1 | -355 | 106 | L8; F12; W37; I52; L55; S56; M61; I62; A73; A77; H80; L90; V94; K110; T111; I114; A115; F118 |

**Table 3.6** **Moth PBPs: regions of high TIE between the CMET probe and the proteins**

There are two distinct and contiguous CMET clusters of nearly equal volume and TIE in *Bmor*PBP1 as compared to only one cluster of much higher TIE and volume in *Msex*PBP1. The *Msex*PBP1 cluster overlaps to a large extent with the 2nd ranking cluster of *Bmor*PBP1 as shown in Figure 3.11.

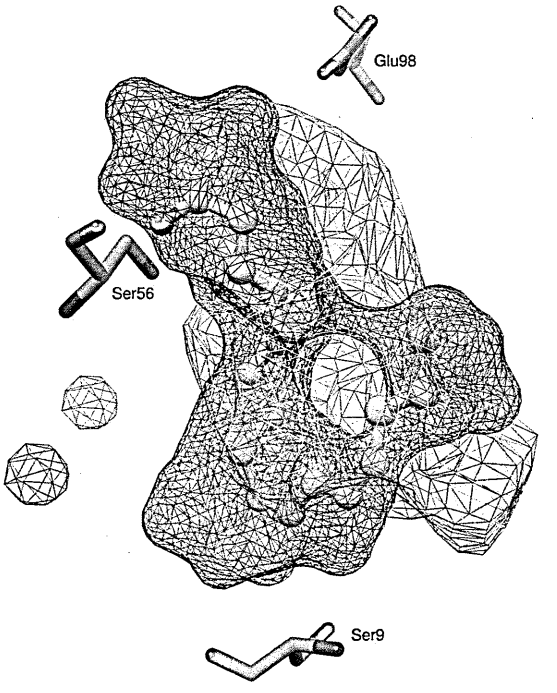



**Figure 3.11** **Moth PBPs: TIE clusters between the proteins and the CMET probe**

The X-ray protein models have been superimposed. The *Bmor*PBP1 energy clusters are depicted red (highest TIE) and blue (lower TIE). The *Msex*PBP1 highest energy cluster is shown in yellow. A low energy outlier is shown in green

The results of probing the binding sites of *Bmor*PBP1 and *Msex*PBP1 with the OA and CMET probes do not reveal any major differences between the two proteins. For mainly hydrophilic ligands with a polar 'head' moiety, it would, most probably, be the type and size of the hydrophilic part of the ligand that would determine its relative orientation in the binding cavity through its interactions with the side chains of the surrounding residues.

### 3.3.3.2 *B. mori* and *M. sexta* GOBPs

The highest ranking OA-TIE clusters between the hydroxyl probe and surrounding binding site residues show clear separation in the *Bmor*GOBP2, contrary to *Msex*GOBP2 in which they are contiguous. In both proteins, these clusters are of much smaller volume and, therefore, of considerably lower TIEs than the corresponding clusters of the two PBPs. However, in both proteins they are surrounded by the residues that have been shown to form H-bonds in docking

simulations, namely, Thr9, Trp37, Ser56, Arg67, Glu98, Arg110. The residues within 4 Å of each cluster are shown in Table 3.7 and the relative positions of the clusters in Figure 3.12.

| | Cluster rank | TIE (kcal mol-1) | Volume (Å3) | Cluster surrounding residues |
|---|---|---|---|---|
| *Bmor*GOBP2 | 1 | -65 | 21 | M5; V8; T9; F12; F33; W37; L61; M73; Y76; M90; I94; F118 |
| | 2 | -29 | 8 | V66; R67; E98; R110; V111; V114 |
| *Msex*GOBP2 | 1 | -112 | 31 | I52; I53; S56; L61; L62; T66; R67;I68; M73; I94; E98; R110; V111; V114 |
| | 2 | -111 | 34 | T9; F12; F33; F36; W37; I52; S56; M90; V114; A115; F118; K119 |

**Table 3.7** **Moth GOBPs: regions of high TIE between the OA probe and the proteins**

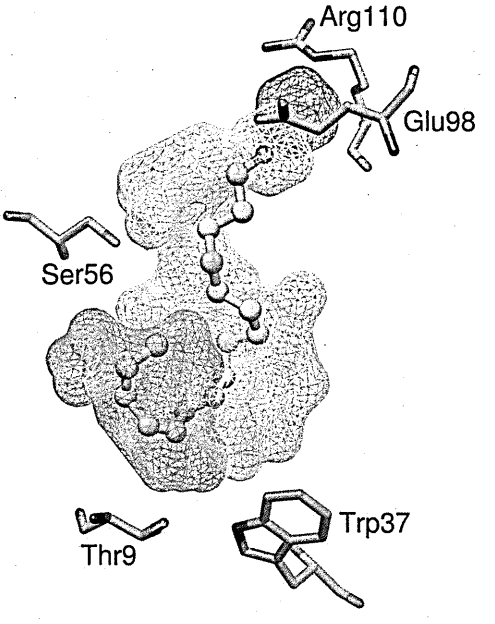



**Figure 3.12** **Moth GOBPs: TIE clusters between the proteins and the OA probe**

The X-ray protein models have been superimposed. The *Bmor*GOBP2 energy clusters are depicted red (highest TIE) and blue (lower TIE). The *Msex*GOBP2 highest energy clusters are shown in yellow. A lower energy outlier is shown in green

On the basis of the above, it would be reasonable to predict that, in the case of *Bmor*GOBP2, small polar ligands would have a tendency to bind preferentially in the region of the highest TIE cluster (Cluster rank 1, Table 3.7).

The results of probing the binding sites with the methyl-carbon probe (CMET) are summarised in Table 3.8. There are two distinct contiguous CMET clusters in *Msex*GOBP2 of almost equal TIE. One of them partly overlaps with the only cluster of *Bmor*GOBP2 of any significance, which is located close to Thr9, Trp37 and three conserved phenylalanine residues (Phe12, Phe33, Phe36).

| | Cluster rank | TIE (kcal mol-1) | Volume (Å3) | Cluster surrounding residues |
|---|---|---|---|---|
| *Bmor*GOBP2 | 1 | -57 | 18 | M5; V8; T9; F12; F33; F36; W37; L61; M73; Y76; M90; F118 |
| | 2 | -14 | 4 | M62; V66; R67; M68; E98; R110; V111; V114 |
| *Msex*GOBP2 | 1 | -118 | 30 | I52; S56; L61; L62; T66; R67; I68; M73; I94; E98; R110; V111; V114 |
| | 2 | -114 | 32 | M5; V8; T9; F12; F33; F36; W37; I52; S56; M90; I94; A115; F118 |

**Table 3.8** **Moth GOBP2: regions of high TIE between the CMET probe and the proteins**

The regions of favourable TIE of the CMET probe for one protein relative to the second one are shown in Figure 3.13.

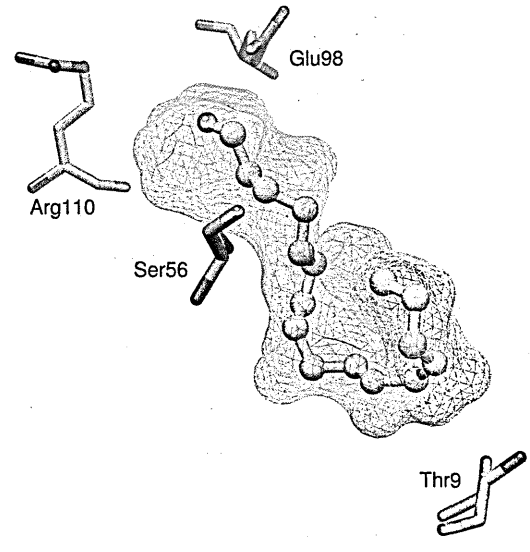



**Figure 3.13** **Moth GOBPs: TIE clusters between the proteins and the CMET probe**

The X-ray protein models have been superimposed. The *Msex*GOBP2 energy clusters are depicted red (highest TIE) and blue (lower TIE). The *Bmor*GOBP2 highest energy cluster is shown in yellow

The TIEs for the OA and CMET probes are much higher for the PBPs than the corresponding GOBPs. This would indicate that the binding cavities of these proteins can probably accommodate a greater number of ligand conformations of the same or nearly the same binding energies as compared to the corresponding GOBPs. Other than this, the results obtained on probing the binding sites of the *B. mori* and *M. sexta* OBPs do not reveal any striking differences in interaction energies between the proteins and the CMET or OA probes. In any given OBP there are distinct regions in its binding cavity that are favourable in terms of binding energy interactions from its orthologous partner. Screening of small ligand libraries to identify compounds docked in these distinct regions would be an appropriate approach to achieve fragment-based design of ligands with a potential to disrupt olfactory function.

As mentioned earlier, OBPs have been shown experimentally to be promiscuous in regard to ligand binding. Binding affinities are dominated by hydrophobic interactions and ligand selectivity may depend on a very fine interplay of electrostatic interactions and weak interactions such as π-π, amino-aromatic interactions, etc. The binding cavity of the *B. mori* and *M. sexta* OBPs is rich in amino acids with aromatic side chains which can play an important part in such interactions.

The promiscuity in ligand binding shown by these OBPs coupled with some selectivity for molecules with specific geometries could well be an evolutionary adaptation supporting the dual "carrier-scavenger" model of OBP function.

# 3.4 MD simulations

## 3.4.1 Binding of bombykol

### 3.4.1.1 Properties of the molecular systems used

The precise protocol for the MD simulations is described in Section 2.2.3.1. The alpha chains of the Protein Data Bank structures 1dqe and 2wch were taken as the starting structures for the MD simulations of *Bmor*PBP1 and *Bmor*GOBP2, respectively. The homology models *Msex*PBP1 and *Msex*GOBP2, prepared as described in Section 3.2.1.2, were used for the MD simulations of *M. sexta* proteins. MD production simulations were carried out in explicit water with the particle mesh Ewald (PME) method after an initial 2ns of equilibration. Table 3.9 provides a summary of the properties of the molecular systems used in MD simulations. For the *B. mori* models, the initial conformations of bombykol and bombykal were those of the corresponding crystallographic models. For the *M. sexta* simulations, the initial bombykol and bombykal conformations were taken from the docking simulations.

| Complex | PDB id | No. aa | No. atoms | No. counterions | No. HOH | gyration radius (Å) | Box (Å) |
|---|---|---|---|---|---|---|---|
| *Bmor*PBP1-BOM | 1dqe | 137 | 2099 | 8 | 4033 | 14.0 | 69x69x69 |
| *Bmor*GOBP2-BOM | 2wch | 141 | 2206 | 13 | 4260 | 14.2 | 70x70x70 |
| *Msex*PBP1-BOM | hm | 142 | 2172 | 9 | 4240 | 14.6 | 70x70x70 |
| *Msex*GOBP2-BOM | hm | 141 | 2235 | 11 | 4334 | 14.1 | 70x70x70 |

**Table 3.9 Bombykol: Properties of molecular systems used in MD simulations**

No. aa & No. atoms are the number of residues and number of atoms of the protein; No. counterions is the number of $Na^+$ ions used to neutralise the system; No. HOH is the number of water molecules contained in the solvation box; hm stands for homology model

### 3.4.1.2 Time-dependent properties

For each trajectory, the time-series of the RMSD backbone atoms from the respective experimental structures is given in Table 3.10. For all systems with the exception of *Msex*PBP1, the backbone RMSD values with respect to the starting structures were in the range of 0.7-1.6 Å. The larger RMSD observed for *Msex*PBP1 is mainly due to backbone fluctuations of the unstructured carboxyl tail of the protein (residues 123-142), which is considerably longer than of *Bmor*PBP1 (see Figure 3.1). This is not the case for *Bmor*GOBP2 and *Msex*GOBP2 in which the carboxyl tail ends are of almost equal length and form a helix.

| Ref. structure[1] | complex | residues | avg[2] (Å) | sd[3] | min[4] (Å) | max[4] (Å) | num[5] |
|---|---|---|---|---|---|---|---|
| 1dqe | *Bmor*PBP1 - BOM | 1-137 | 1.2 | 0.12 | 0.9 | 1.6 | 1600 |
| hm | *Msex*PBP1 - BOM | 1-142 | 2.4 | 0.21 | 1.9 | 2.9 | 1000 |
| | | 1-125 | 1.8 | 0.17 | 1.4 | 2.3 | 1000 |
| 2wc6 | *Bmor*GOBP2-BOM | 1-141 | 1.1 | 0.16 | 0.7 | 1.5 | 1000 |
| hm | *Msex*GOBP2-BOM | 1-141 | 0.9 | 0.08 | 0.7 | 1.3 | 1000 |

**Table 3.10** **Bombykol: Backbone RMS deviation of MD simulation models**

1. hm denotes homology model; 2. average backbone RMSD during the trajectory with respect to the reference structure; 3. standard deviation; 4. minimum and maximum RMSD with respect to the reference structure; 5. number of frames

The RMSD of the backbone atoms of the proteins and the bombykol (BOM) are shown in Figure 3.14. The figures also show hydrogen bond formation between BOM and the respective proteins. Binding of BOM to *BmorP*BP1 and *Msex*PBP1 shows that the ligand underwent major conformational transitions as evidenced by the abrupt fluctuations of the ligand RMSD. This is in line with the X-ray model of *Bmor*PBP1 which shows significantly larger B-factors for bombykol compared to the average backbone B-factors suggesting that bombykol can attain a large number of conformations within the binding pocket of the protein. The RMSD fluctuations of BOM during the MD trajectories of *Bmor*GOBP2 and *Msex*GOBP2

are less pronounced. This may be due to the much higher H-bond occupancy that is observed between these proteins and bombykol in comparison to the corresponding PBPs. Bombykol forms H-bonds simultaneously with Glu98 and Arg110 of the proteins, which would anchor the ligand and reduce its degrees of freedom. H-bond formation during the MD production trajectories is shown in Table 3.11.

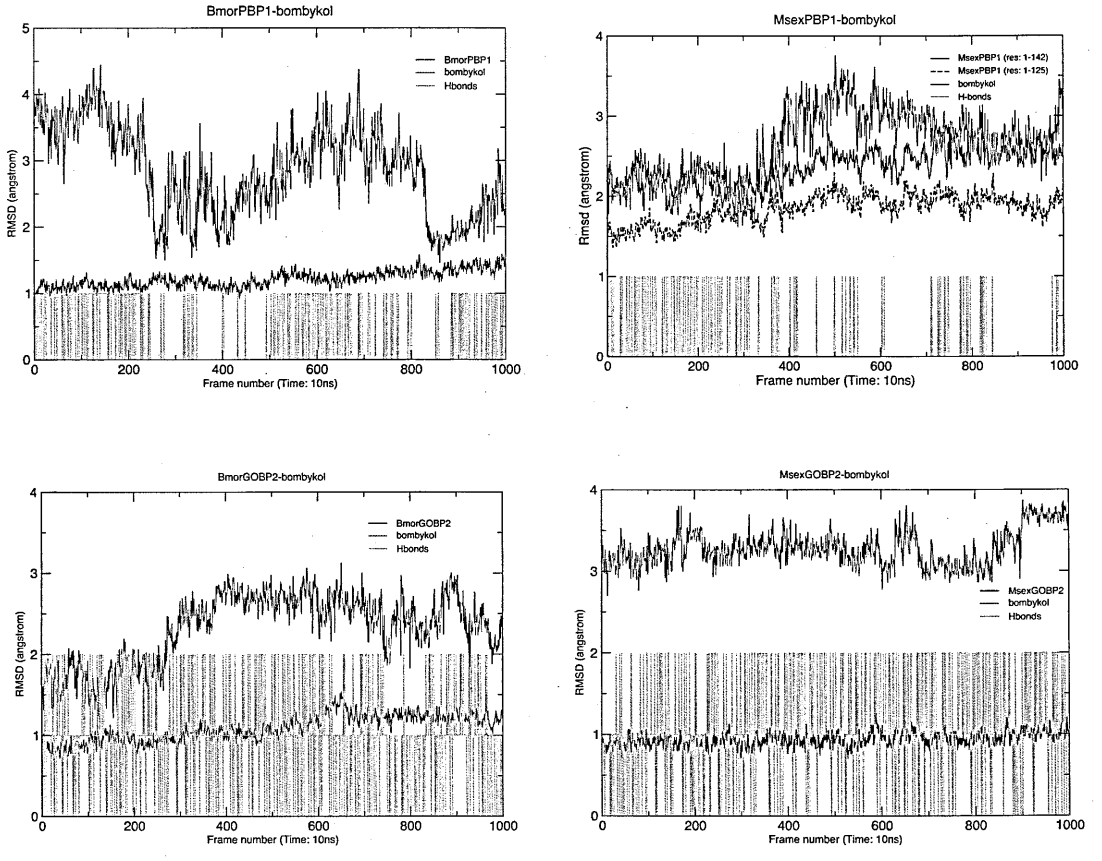



**Figure 3.14 RMSD time series of OBP backbone atoms and bombykol from the starting structures**

The equilibration phase is not included

In both *B. mori* complexes, the hydrogen bonds formed between the ligand and the proteins are the same as those observed in the corresponding X-ray models [58] [64]. In the case of *Bmor*PBP1 the oxygen atom of the hydroxyl moiety of bombykol serves as electron donor to the -OH proton of Ser56, whereas the oxygen atom of the latter is shown to act transiently as electron donor to the -OH

proton of bombykol. Hydrogen bond formation between bombykol and Ser56 was observed in the 2nd half of the 10ns trajectory while the first quarter of the 10ns trajectory was dominated by hydrogen bonds between bombykol and Glu98 mainly and between bombykol and Met61 and Thr111, transiently. These observations are in agreement with the findings reported by Charlier *et al.* [97] who conducted *Bmor*PBP1-bombykol MD simulations over 10ns and Grubmüller *et al.* [98] whose simulations were much longer (45-60ns).

| | VMD | | | AMBER (cpptraj) | | |
|---|---|---|---|---|---|---|
| **protein** | **donor** | **acceptor** | **%** | **donor** | **acceptor** | **%** |
| *BmorPBP1* | BOM1@OH | GLU98@OE | 17 | SER56@HG | BOM@O1 | 22 |
| | BOM@OH | SER56@OG | 1 | BOM@OH | GLU98@OE | 21 |
| | SER56@HG | BOM@O1 | 14 | BOM@OH | SER56@OG | 2 |
| | | | | BOM@OH | THR111@O | 2 |
| *MsexPBP1* | BOM@OH | GLY66@O | 1 | BOM@OH | SER56@OG | 15 |
| | BOM@OH | SER56@OG | 9 | TRP37@NHE1 | BOM@O | 14 |
| | BOM143@OH | PHE33@O | 2 | BOM@OH | PHE33@O | 3 |
| | TRP37@NHE1 | BOM143@O | 4 | BOM@OH | GLY66@O | 2 |
| | | | | BOM@OH | WAT@O | 3 |
| | | | | BOM@OH | CYS9@O | 1 |
| *BmorGOBP2* | BOM@OH | GLU98@OE2 | 66 | ARG110@HE | BOM@O | 91 |
| | ARG110@HE | BOM@O1 | 25 | BOM@OH | GLU98@OE2 | 89 |
| | BOM@OH | VAL66@O | 2 | BOM@OH | VAL66@O | 6 |
| *MsexGOBP2* | BOM@OH | GLU98@OE | 73 | BOM@OH | GLU98@OE | 99 |
| | ILE68@NH | BOM@O | 1 | ARG110@NH | BOM@O | 67 |
| | ARG110@NH | BOM@O | 36 | ILE68@NH | BOM@O | 2 |

**Table 3.11** **H-bond formation between the *B. mori* and *M. sexta* OBPs and bombykol**

% = H-bond occupancy; @ is atom name; H-bonds of occupancy less that 1% are not shown; A definition of occupancy and H-bond distance and angle cutoffs are given in Section 2.2.6.2

In the case of *Bmor*GOBP2, hydrogen bond formation between bombykol and the protein occurs throughout the trajectory. The hydroxyl oxygen atom of bombykol serves as proton acceptor forming hydrogen bonds with the guanidinium

hydrogens of Arg110. Hydrogen bond formation also occurs between one of the oxygens of Glu98 and -OH of bombykol. Thus, it appears that the triangular hydrogen bonding between Glu98---BOM---Arg110 takes place simultaneously. A similar network of H-bonding interactions is seen in the X-ray model which shows BOM in H-bonding distance with Glu98 and a water molecule. The latter is within H-bonding distance from Arg110. H-bonding interactions of the *M. sexta* proteins with bombykol are very similar to the ones observed with the corresponding *B. mori* proteins. In *Msex*GOBP2, the hydroxyl group of bombykol acts both as proton donor to the side chain carboxyl oxygens of Glu98 and as an acceptor receiving protons from the side-chain guanidinium group of Arg110. In *Msex*PBP1, H-bond formation of low occupancy between bombykol and Ser56 and bombykol and Tpr37 was observed indicating a flip in the orientation of the ligand during the trajectory. Figure 3.15 shows fluctuations in the interatomic distances between the hydrogen atom of indole of Trp37 and the hydroxyl oxygen atoms of bombykol and Ser56.

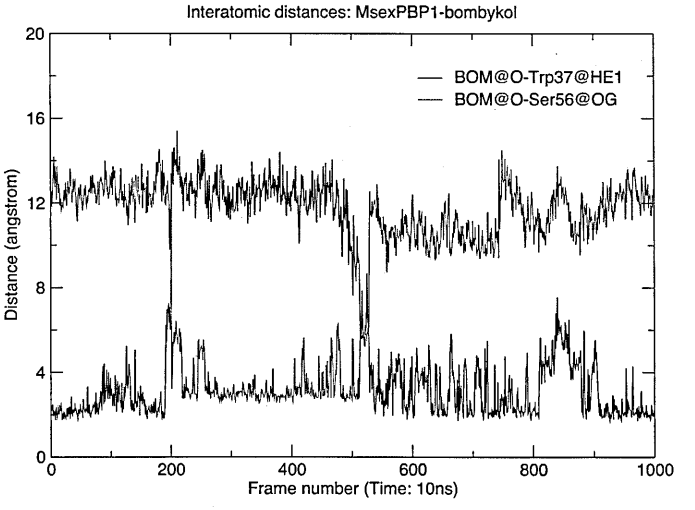



**Figure 3.15** ***Msex*PBP1: Interatomic distances between the hydrogen atom of the indole moiety of Trp37 and the hydroxyl oxygen atoms of bombykol and Ser56**

The abrupt shift in the orientation of the polar head of bombykol is apparent. A longer simulation would be instructive in establishing whether this change in the orientation of the ligand occurs periodically.

#### 3.4.1.3 Ligand entry and exit pathways into and out of BmorGOBP2

Two main models for receptor activation by OBPs were discussed earlier (Section 1.3.3). One of these proposes that upon release of the odorant molecule in the vicinity of the odorant receptor, the odorant induces conformational changes in receptor or co-receptor proteins (SNMP) [42] which, in turn, induce a cascade of biochemical reactions leading to odorant recognition. The second hypothesis suggests that the activation of olfactory response is due to the direct interaction between the OBP-ligand complex and the olfactory receptor. MD studies with the *Bmor*PBP1-bombykol complex have led to the conclusion that there are two possible gates for bombykol to enter and exit from the binding pocket [98, 189]. It was suggested that the front gate involves residues 60-69 forming a flap that is controlled by His69, His70 and His95. The latter are conserved throughout the moth PBPs and GOBPs. The rear gate is formed by the terminal fractions of the protein (residues 1-14 and 126-137).

There have been no MD or other studies regarding the entry or exit of bombykol in and out of *Bmor*GOBP2, although it has been suggested that the binding site of *Bmor*GOBP2 may be gated by the loop spanning residues 60-69 similarly to the front gate of *Bmor*PBP1. There is no rear gate to the binding pocket of *Bmor*GOBP2 as the opening that would correspond to the rear gate of *Bmor*PBP1 is occluded by a bulge formed by residues 33-35. An alternative model of activation has been suggested involving the salt-bridge between Lys89 and Glu125, which is structurally analogous to the one in LUSH formed by residues

Lys87 and Asp118. In LUSH, the disruption of this salt-bridge results in a conformational change in the LUSH-pheromone complex that constitutively activates the T1 neurons in the presence of the olfactory receptor [64].

To test these hypotheses, the stability of the salt-bridge between Lys89 and Glu125 was analysed over the 10ns *Bmor*GOBP2-bombykol MD trajectory. It was shown (Figure 3.16) that the salt-bridge between Lys89 and Glu125 is broken for a large part of the trajectory. However, it is not possible to conclude with certainty whether binding of bombykol induces a conformational change analogous to the one observed in the LUSH-pheromone complex and a longer MD simulation would be required to test the proposed hypothesis.

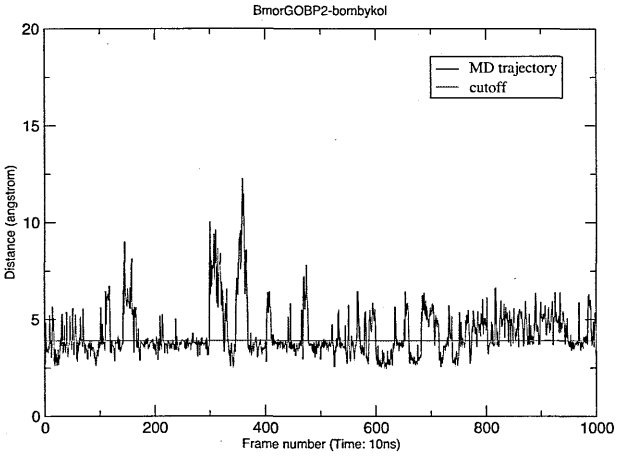



**Figure 3.16** **Salt bridge formation and disruption between residues E125 and K89**

The distance between the N and O atoms of these residues in the X-ray model is taken as the cutoff distance

Additionally, comparison of the B-factors of the protein Cα atoms from the 10ns MD trajectory of the *Bmor*GOBP2-bombykol complex with the corresponding crystallographic ones. Figure 3.17 depicts a comparative plot of the B-factors showing a distinct peak in the trajectory, a major one at residue 67 and, a smaller one, at residue 125 indicative of higher atomic positional fluctuations in the vicinity

of these residues. The amplitude of rest of the peaks from the MD simulation and the X-ray model are of the same range indicating that the 10ns trajectory provided adequate sampling to compare to the crystal structure. As will be seen immediately below, residue 67 (Arg) may be part of the entry/exit gate of the protein.

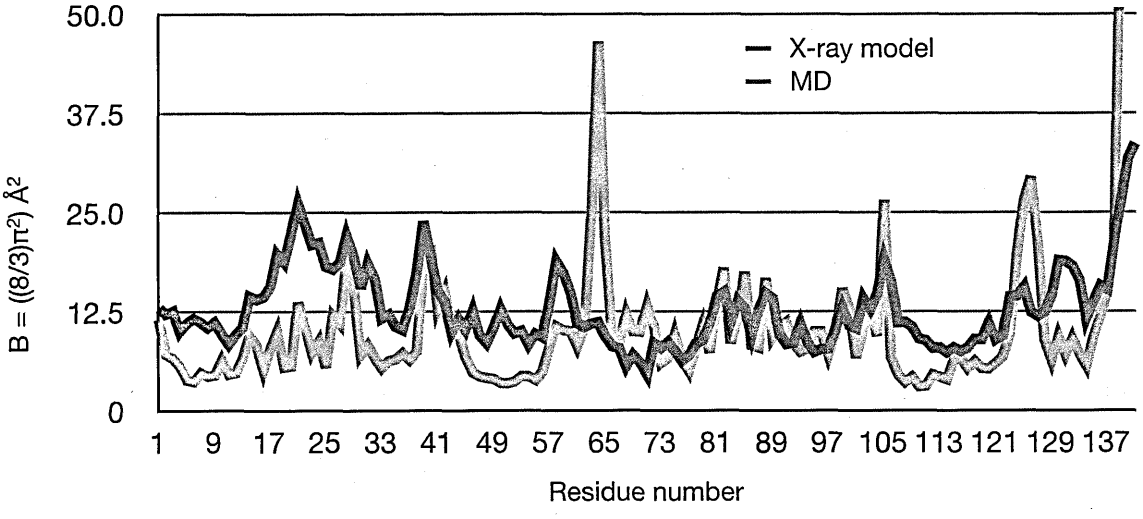



**Figure 3.17** **B-factors of Cα atoms of *Bmor*GOBP2 trajectory and X-ray model** [10]

SMD simulations were used as an alternative approach to identify possible exit pathways. The technique is described in Section 2.1.9. The starting coordinates of the complex were taken after 10ns of production MD. The initial restrained distance was defined by the centres of mass of bombykol and the residues lying within a distance of 4 Å from it (~1.5 Å). This was increased to 14.5 Å by applying a force constant of 500 kcal/mol-Å$^2$ and changing the spring length at constant 1.75 Å ns$^{-1}$ velocity. The process was repeated three times by applying the same force constant at constant velocity of 0.75 Å ns$^{-1}$. In all three SMD runs bombykol was released through a channel gated by Arg67 and Glu98 (Supplementary Material: SMD: *Bmor*GOBP2, exit pathways). The presence of this channel was also detected by the MolAxis program described in Section 2.2.7. The program

[10] See page 84 for a definition of B-factor units

identified several channels emanating from the binding site of *Bmor*GOBP2. Residues Arg67 and Glu98 constitute the narrowest part of the top ranked channel.

Hydrogen bond analysis of the SMD trajectory shows the gradual breaking of bonds between bombykol Glu98 and Asp110 and the formation of H-bonds with the hydroxyl oxygen as proton donor and the side chain oxygen (OD) of Asp102 as acceptor with an occupancy of 16.75%. The bombykol hydroxyl oxygen is a proton acceptor of the side chain nitrogens (NE, NH) of Arg67 which act as proton donors (occupancy 2.38%). Transient formation of H-bonds was also observed between bombykol and Thr103. The opening of the exit channel gate involved the breaking of the salt-bridge between Arg67 and Asp102. Figure 3.18 shows the channel identified by MolAxis and the exit pathway of bombykol as seen from the SMD trajectory.

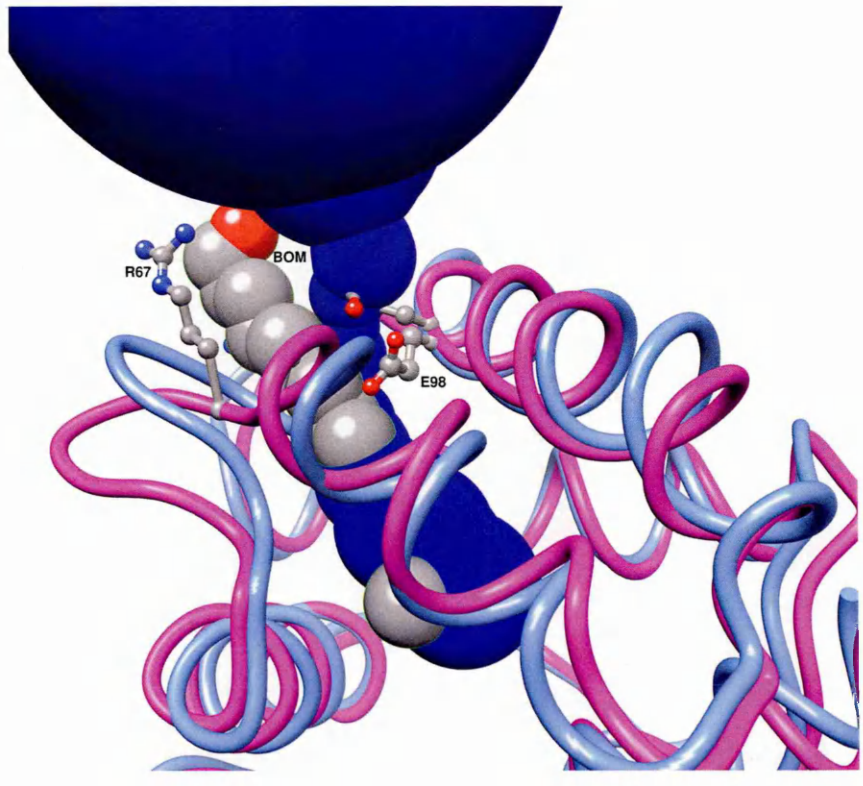



**Figure 3.18** **Dissociation pathway of bombykol from *Bmor*GOBP2**

Blue: channel leading from the binding cavity to the bulk solvent as identified by MolAxis

## 3.4.2 Binding of bombykal

### 3.4.2.1 Properties of the molecular systems used

Protein preparation and MD production simulations were performed as described in Section 2.2.3.1. Table 3.12 provides a summary of the properties of the molecular systems used in MD simulations. For the *Bmor*GOBP2, bombykal (BAL) was used in its crystallographic conformation. For the other models, the initial bombykal conformations were taken from docking simulations that gave the lowest free energies of binding. In two cases, in which the results of the docking simulations were ambiguous, two separate MD simulations were performed. In these simulations the starting conformations of bombykal represented distinct orientations in the binding pocket as discussed in Section 3.3.2. The different simulations for the same complex are distinguished by the subscripts ($BAL_a$ and $BAL_b$).

These simulations were conducted to test whether the starting structure of the ligand has an influence on the time-evolved and time-independent properties of the MD simulations in the time-scale used.

| Complex | PDB id | No. aa | No. atoms | No. counterions | No. HOH | Box (Å) |
|---|---|---|---|---|---|---|
| *Bmor*PBP1-$BAL_a$ | 1dqe | 137 | 2099 | 8 | 4113 | 69x69x69 |
| *Bmor*PBP1-$BAL_b$ | 1dqe | 137 | 2099 | 8 | 4011 | 69x69x69 |
| *Bmor*GOBP2 | 2wch | 141 | 2206 | 13 | 5108 | 74x74x74 |
| *Msex*PBP1 | hm | 142 | 2172 | 9 | 4240 | 70x70x70 |
| *Msex*GOBP2-$BAL_a$ | hm | 141 | 2235 | 11 | 4341 | 70x70x70 |
| *Msex*GOBP2-$BAL_a$ | hm | 141 | 2235 | 11 | 4363 | 70x70x70 |

**Table 3.12** **Bombykal: Properties of molecular systems used in MD simulations**

No. aa & No. atoms are the number of residues and number of atoms of the protein; No. counterions is the number of $Na^+$ ions used to neutralise the system; No. HOH is the number of water molecules contained in the solvation box; hm stands for homology model. Radii of gyration are reported in Table 3.7

### 3.4.2.2 Time-dependent properties

For each trajectory, the time-series of the RMSD backbone atoms from the respective experimental structures is given in Table 3.13. For all systems with the exception of MsexPBP1, the backbone RMSD values, with respect to the starting structures, were in the range of 0.7-1.7 Å. The larger RMSD observed for *Msex*PBP1 is due mainly to backbone fluctuations of the unstructured carboxyl tail of the protein (residues 123-142) as observed in the case of the *Msex*PBP1-bombykol complex (see Section 3.4.1.2).

| Ref. structure[1] | complex | residues | avg[2] (Å) | sd[3] | min[4] (Å) | max[4] (Å) | num[5] |
|---|---|---|---|---|---|---|---|
| 1dqe | *Bmor*PBP1- $BAL_a$ | 1-137 | 1.3 | 0.127 | 1.0 | 1.7 | 1000 |
| 1dqe | *Bmor*PBP1- $BAL_b$ | 1-137 | 1.3 | 0.159 | 0.9 | 1.7 | 1000 |
| hm | *Msex*PBP1 - BAL | 1-142 | 2.2 | 0.168 | 1.7 | 2.6 | 1000 |
| | | 1-125 | 1.8 | 0.110 | 1.5 | 2.2 | 1000 |
| 2wch | *Bmor*GOBP2 - BAL | 1-141 | 0.9 | 0.093 | 0.7 | 1.3 | 1000 |
| hm | *Msex*GOBP2 - $BAL_a$ | 1-141 | 1.2 | 0.107 | 0.9 | 1.6 | 1000 |
| hm | *Msex*GOBP2 - $BAL_b$ | 1-141 | 1.0 | 0.088 | 0.8 | 1.3 | 1000 |

**Table 3.13** **Bombykal: Backbone RMS deviation of MD simulation models**

1. hm denotes homology model; 2. average backbone RMSD during the trajectory with respect to the reference structure; 3. standard deviation; 4. minimum and maximum RMSD with respect to the reference structure; 5. number of frames

In the case of *Bmor*PBP1 and *Msex*GOBP2, two separate MD simulations were performed. In these the starting coordinates of the proteins were identical whereas the starting coordinates of bombykal were different. These are indicated with the subscripts a and b, respectively. In *Bmor*PBP1-$BAL_b$, the starting position of bombykal is that of the X-ray model. In the other three MD simulations (*Bmor*PBP1-$BAL_b$, *Msex*GOBP-$BAL_{a/b}$), the initial coordinates of bombykal were taken from docking simulations.

The RMSD of BAL and of the backbone atoms of the proteins are shown in Figure 3.19. The figure also shows hydrogen bond formation between BAL and the respective proteins. It also shows that all proteins are equilibrated throughout the production trajectories.

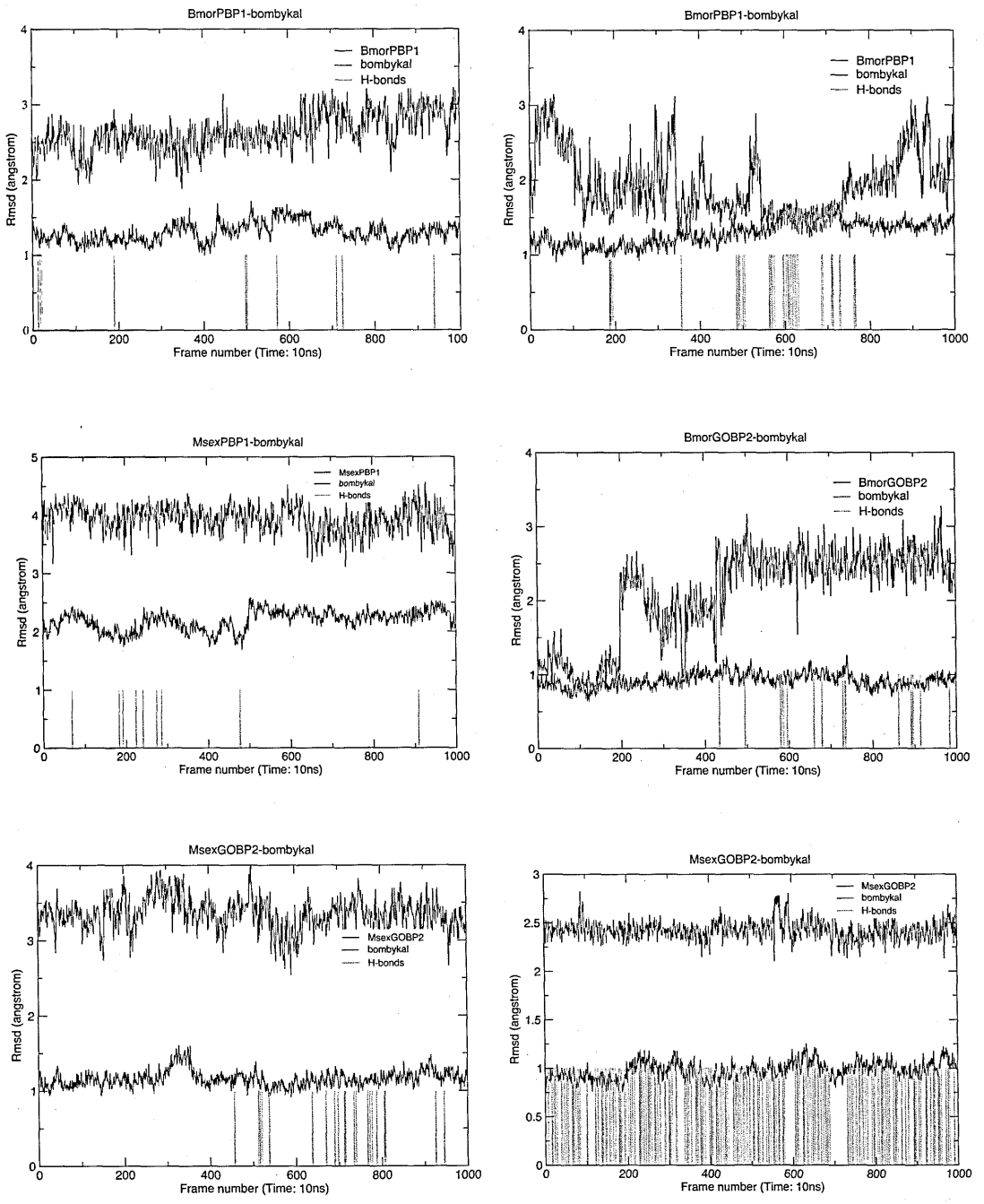



**Figure 3.19** **RMSD time series of OBP backbone atoms and bombykal from the starting structures**

In two of the three *B. mori* OBP trajectories, the RMSD of BAL shows abrupt

transitions indicative of major conformational changes during the MD trajectory. It can also be seen that hydrogen bond interactions between BAL and the receptors are less frequent in comparison to the respective receptor-BOL complexes. Hydrogen bond interactions are summarised in Table 3.14.

| | VMD | | | AMBER (cpptraj) | | |
|---|---|---|---|---|---|---|
| **protein** | **donor** | **acceptor** | **%** | **donor** | **acceptor** | **%** |
| *Bmor*PBP1.a | TRP37@NHE1 | BAL@O | 1 | TRP37@NHE1 | BAL@O | 10 |
| | | | | WAT@OH | BAL@O | 1 |
| *Bmor*PBP1.b | SER56@HG | BAL@O | 5 | SER56@HG | BAL@O | 10 |
| *Bmor*GOBP2 | WAT@OH | BAL@O | 19 | WAT@OH | BAL@O | 33 |
| | ARG110@NH | BAL@O | 2 | ARG110@NH | BAL@O | 11 |
| *Msex*PBP1 | | | | THR111@HG | BAL@O | 1 |
| | | | | WAT@OH | BAL@O | 1 |
| | | | | LYS110@NZH | BAL@O | 1 |
| *Msex*GOBP2.a | THR9@HG1 | BAL@O | 2 | WAT@OH | BAL@O | 5 |
| | | | | THR_9@HG | BAL@O | 5 |
| | | | | TRP37@NHE | BAL@O | 1 |
| *Msex*GOBP2.b | ARG110@NH | BAL142@O1 | 27 | ARG110@NH | BAL@O | 58 |

**Table 3.14** **H-bond formation between the *B. mori* and *M. sexta* OBPs and bombykal**

% = H-bond occupancy; @ is atom name; H-bonds of occupancy less that 1% are not shown; A definition of occupancy and H-bond distance and angle cutoffs are given in Section 2.2.6.2

The carbonyl group of bombykal can only be a proton acceptor. Thus hydrogen bond interactions between bombykal and the moth OBP/PBPs exclude anionic side chains such as the carboxylate groups of Glu98. H-bond formation between

*Bmor*PBP1 and bombykal is transient over the two MD simulations and dependent on the starting conformation of the ligand. Figure 3.20 depicts the starting coordinates of the bombykal.

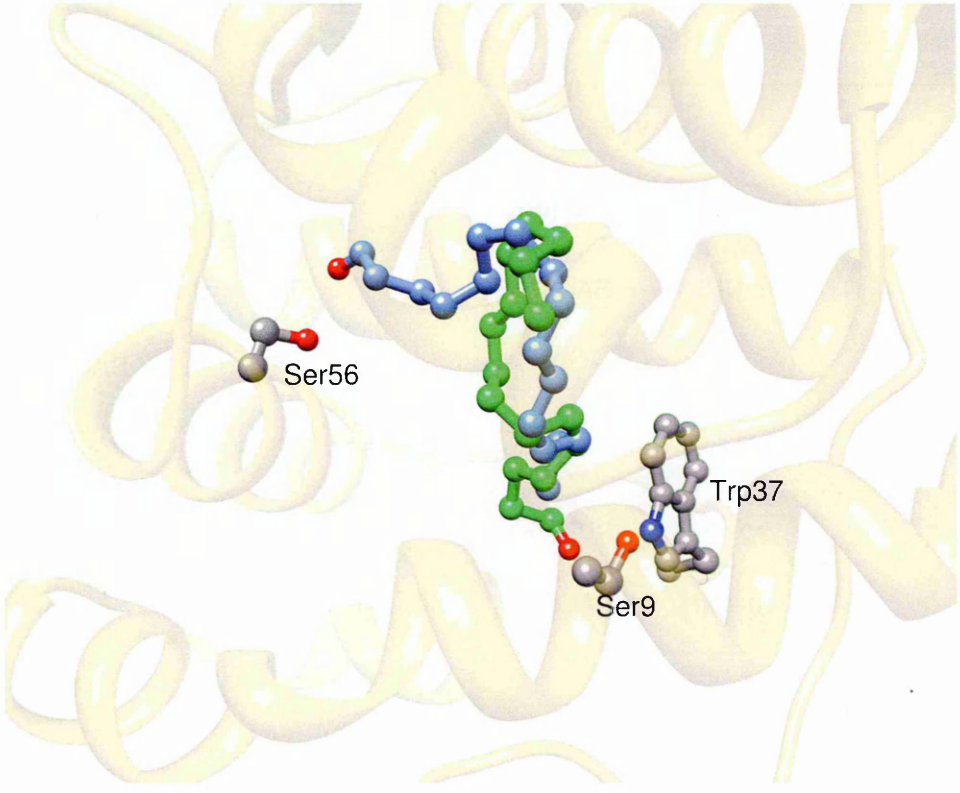



**Figure 3.20** **Starting coordinates of bombykal in MD simulations with *Bmor*PBP1 and *Msex*GOBP2**

In the case of the *Bmor*PBP1-BAL$_{a/b}$ and *Msex*GOBP2-BAL$_{a/b}$ trajectories, bombykal formed H-bonds with residues that were in its vicinity at the start of the simulation time as shown in Figure 3.20. Ser9 and Trp37 have been shown to play a role in the dissociation pathway of bombykol from *Bmor*PBP1, with Ser9 forming hydrogen bonds transiently [189]. In the course of the respective simulations, the orientation of bombykal did not change with respect to its starting positions. It is likely that the short nanosecond timescale of conventional MD simulations of these complexes is inadequate to overcome larger energy barriers and sample regions of the configurational space that are also sampled at physiological conditions. A much longer trajectory or a different type of MD approach (e.g. simulated

annealing ([102], [103]) would be required to overcome such barriers. In the one MD simulation in which bombykal formed H-bonds with the indole -NH of the conserved TRP37, H-bonding to water and, transiently, to Ser9 was also observed suggesting the possible formation of a hydrogen bond network between bombykal, water, Trp37 and Ser9. The effect of this on ligand stabilisation in this conformation is further discussed in Section 3.5.1.

Similar observations were made in the two separate simulations involving *Msex*GOBP2 and bombykal. In one of them, H-bond formation of high occupancy involved one of the guanidinium hydrogens of Arg110. In the second, transient H-bond formation between bombykal, Trp37, Thr9 and water was observed. *Bmor*GOBP2 formed H-bonds with Arg110, albeit of lower frequency. This observation is consistent with the crystallographic model (PDB id: 2wch) which shows a bridged hydrogen bond between the aldehyde moiety of bombykal and a water molecule, which, in turn, forms a hydrogen bond with one of the guanidinium hydrogens of Arg110. No significant H-bond formation was observed between *Msex*PBP1 and bombykal during the time of the MD simulation.

## 3.4.3 Time-independent properties

### 3.4.3.1 Binding free energy components

Estimates of the binding free energies of the *B. mori* and *M. sexta* complexes with bombykol and bombykal were obtained by calculating the interaction and solvation free energies of the complexes, receptors and ligands. The calculations were performed at 0.1 mM ionic strength.

Molecular Mechanics Poisson Boltzmann or Generalised Born Surface Area (MM-PB(GB)SA calculations were performed as described in Section 2.2.5. It was applied on the ensemble of conformations generated from production time MD simulations to calculate the free energy change between the bound and free states of the receptors and ligands at hand. Although neither method (GB or PB) is suitable to predict absolute free energies, they are accurate enough to compare relative free energies of binding of different ligands to the same receptor [101]. In high-throughput virtual screening studies, MM-PBSA and MM-GBSA have been shown to discriminate true binders from a much larger number of decoys, and give binding free energies in excellent agreement with experimental data [190]. In a more recent study evaluating the binding affinity of a set of 156 ligand to seven protein families, the statistical correlation between binding energies calculated by means of MM-PB(GB)SA models and experimental free energies was found to be in the range of 0.56 - 0.79 across all protein families [191]. Yet, in binding affinity calculations involving 59 ligands and six protein systems, good correlations with experimental results ($r2 > 0.5$) were found in three out of the six systems studied. [192]. In the light of these studies, it has been concluded that the accuracy of implicit solvation methods to predict binding affinities in agreement with experimental ones depends on factors, such as the structural similarity of the tested ligands, specific features of protein-ligand interactions, as well as the binding affinity range of the analysed ligand data sets [101].

MM-GBSA has been selected as the method of choice for the evaluation of binding affinities reported in this work as it has been found to outperform MM-PBSA in both binding pose predictions and binding free-energy estimations [192]. MM-PBSA results can be found in the Supplementary Materials CD in the MMPBSA subdirectories of each of the simulations performed. These are in overall

agreement with the corresponding MM-GBSA, although as expected, the actual values are different. The binding free energies are decomposed to contributions of gas-phase energies, solvation free energies, as well as entropic contributions. The free-energies of binding are calculated according to:

$$\Delta G_{binding} = \Delta G_{MM-GBSA} + \Delta H_{rot/trans} - T\Delta S \qquad 3.1$$

where $\Delta H_{rot/trans}$ corresponds to the six rotational and translational degrees of freedom and has a value, in the classical limit, of:

$$\Delta H_{rot/trans} = 3RT = 1.8 kcal mol^{-1} \qquad 3.2$$

### *B. mori, M. sexta* OBP - bombykol complexes: Binding free energy components

A comparison of binding free energies of the *B. mori* and *M. sexta* OBPs binding proteins with bombykol (BOM) are shown in Table 3.15. $\Delta G_{total}$, in the table, does not include the $\Delta H_{rot/trans}$ term mentioned above. The MM-GBSA results, predict that *Bmor*GOBP2 and *Msex*GOBP2 have considerably higher binding affinities to bombykol than the corresponding PBPs with $\Delta G_{gas+solv}$ values in the order:

*Msex*GOBP2 > *Bmor*GOBP2 > *Msex*PBP1 > *Bmor*PBP1

In all complexes, van der Waals interactions ($\Delta H_{vdW}$), electrostatic energies ($\Delta H_{elec}$) and the non-polar part of solvation free energy ($\Delta G_{np}$) make the greatest contribution to the binding free energies, whereas the contribution of the polar part of the solvation energy ($\Delta G_{GB}$) on the binding equilibrium does not favour protein-ligand binding. However, it should be noted that the calculation of long range electrostatic interactions is largely dependent on the solvation model used, the ionic strength of the solutions and dielectric constants of solute regions. Notwithstanding the above, the major determinants of the differences observed in

$\Delta G_{gas+solv}$ are electrostatic interactions. $\Delta H_{elec}$ for the *Bmor*GOBP2 and *Msex*GOBP2 complexes are much larger than the corresponding $\Delta H_{elec}$ for the PBPs.

| | BmorPBP1 | | MsexPBP1 | | BmorGOBP2 | | MsexGOBP2 | |
|---|---|---|---|---|---|---|---|---|
| Contrib.[a] | Δ value[b] | σ[c] | Δ value[b] | σ[c] | Δ value[b] | σ[c] | Δ value[b] | σ[c] |
| $\Delta H_{vdW}$ | -41.4 | 2.2 | -43.0 | 2.5 | -42.8 | 2.9 | -41.0 | 2.8 |
| $\Delta H_{elec}$ | -6.4 | 3.6 | -5.4 | 2.9 | -16.5 | 3.7 | -17.8 | 3.0 |
| $\Delta H_{gas}$ | -47.8 | 3.9 | -48.5 | 3.8 | -59.3 | 3.6 | -58.8 | 3.3 |
| $\Delta G_{GB}$ | 17.1 | 1.9 | 16.7 | 2.1 | 22.1 | 1.7 | 20.1 | 1.8 |
| $\Delta G_{np}$ | -5.9 | 0.3 | -6.1 | 0.3 | -6.4 | 0.2 | -6.2 | 0.2 |
| $\Delta G_{solv}$ | 11.2 | 1.8 | 10.6 | 2.0 | 15.6 | 1.7 | 13.9 | 1.7 |
| $\Delta G_{gas+solv}$ | -36.6 | 2.9 | -37.8 | 2.7 | -43.7 | 2.9 | -44.9 | 2.6 |
| $\Delta TS_{total}$ | -21.9 | 2.4 | -18.5 | 3.9 | -21.7 | 1.6 | -19.3 | 2.6 |
| $\Delta G_{total}$ | -14.7 | | -19.4 | | -20.2 | | -23.8 | |

**Table 3.15** **Binding free energies and free energy components of OBP-bombykol complexes**

All values are given in kcal mol$^{-1}$. The standard state is taken to be 1 M. [a] $\Delta H_{elec}$, Coulombic energy; $\Delta H_{vdW}$ van der Waals energy; $\Delta H_{gas} = \Delta H_{elec} + \Delta H_{vdW}$; $\Delta G_{GB}$ polar solvation free energy; $\Delta G_{np}$ non-polar solvation free energy; $\Delta G_{solv} = \Delta G_{GB} + \Delta G_{np}$; $\Delta G_{gas+solv} = \Delta H_{gas} + \Delta G_{solv}$; $\Delta TS_{total}$ total entropy contribution; $\Delta G_{total} = \Delta G_{gas+solv} - \Delta TS_{total}$ ; [b] Average difference (Complex - Receptor - Ligand);
[c] Standard deviation

The total binding free energies include an entropic component. The entropic contribution to the total free energy of binding needs to be treated with caution. Changes in the entropy of solute molecules were determined by normal-mode analysis. The method is not without drawbacks, such as the neglect of anharmonic motions and changes in conformational entropy. Therefore, $\Delta G_{gas+solv}$, also referred to as "effective energy" may be a more reliable comparator for binding affinities than $\Delta G_{total}$.

### *B. mori*, *M. sexta* OBP - bombykal complexes: Binding free energy components

It was mentioned earlier (Section 3.4.2.2) that two separate MD simulations were performed of each of *BmorPBP1* and *Msex*GOBP2. The free energies of binding for each trajectory are given in Table 3.16.

| | *Bmor*PBP1$_{[a]}$ | | *Bmor*PBP $_{[b]}$ | | *Msex*GOBP2 $_{[a]}$ | | *Msex*GOBP2 $_{[b]}$ | |
|---|---|---|---|---|---|---|---|---|
| Contrib.[a] | Δ value[b] | σ[c] | Δ value[b] | σ[c] | Δ value[b] | σ[c] | Δ value[b] | σ[c] |
| $\Delta H_{vdW}$ | -41.6 | 2.0 | -41.5 | 2.6 | -42.6 | 2.7 | -42.7 | 2.2 |
| $\Delta H_{elec}$ | -5.7 | 3.4 | -1.0 | 3.4 | -4.4 | 3.0 | -6.1 | 3.9 |
| $\Delta H_{gas}$ | -47.3 | 3.9 | -42.6 | 4.2 | -46.9 | 4.1 | -48.8 | 4.2 |
| $\Delta G_{GB}$ | 18.2 | 2.8 | 13.8 | 2.8 | 16.6 | 3.0 | 16.0 | 2.7 |
| $\Delta G_{np}$ | -5.8 | 0.2 | -5.9 | 0.3 | -6.1 | 0.3 | -6.0 | 0.2 |
| $\Delta G_{solv}$ | 12.4 | 2.7 | 7.9 | 2.7 | 10.5 | 3.0 | 10.0 | 2.7 |
| $\Delta G_{gas+solv}$ | -34.9 | 2.1 | -34.6 | 2.5 | -36.4 | 3.0 | -38.8 | 2.5 |
| $\Delta TS_{total}$ | -22.2 | 5.4 | -21.5 | 3.4 | -20.5 | 3.1 | -19.3 | 2.1 |
| $\Delta G_{total}$ | -12.7 | | -13.1 | | -15.9 | | -19.5 | |

**Table 3.16** **Binding free energies and free energy components of moth OBP-bombykal complexes**

Notation with regard to superscripts [a], [b] and [c] is given in Table 3.15, p. 138

In the table above, the simulations in which the starting position of bombykal is proximal to Trp37 are designated with the bracketed subscript $_{[a]}$, whereas those in which the starting positions of bombykal are distal to Trp37 are designated with the subscript $_{[b]}$.

For the *Bmor*PBP1-bombykal complexes, $\Delta G_{gas+solv}$ are almost identical (-34.6 and -34.9 kcal mol$^{-1}$, respectively). Energy decomposition shows that the electrostatic energy contribution is much more favourable for *Bmor*PBP1$_{[a]}$ than for *Bmor*PBP1$_{[b]}$ by ~4.8 kcal mol$^{-1}$. However, this is counterbalanced by a

contribution of the polar component of the solvation energy, which is more favourable in the case of *Bmor*PBP1$_{[a]}$ by ~4.5 kcal mol$^{-1}$. To factor in the contribution of the water molecules that were found to form H-bonds with bombykal (Section 3.4.2.2), an additional MM-GBSA calculation was performed on the *Bmor*PBP1$_{[b]}$-BAL complex, in which a single water molecule was added as part of the solute. The calculated value of $\Delta G_{gas+solv}$ -35.9 kcal mol$^{-1}$, the gain of ~1 kcal mol$^{-1}$ being attributable to electrostatic interactions.

It is difficult to ascertain the significance of the presence of water in the binding cavity, although it has been shown that water molecules involved in hydrogen bonds with the ligand and/or protein influence the entropy, enthalpy and heat capacity of binding and can play as important a role as direct interactions. Nonetheless, the effect of water mediated interactions on binding specificity is less clear [193].

A possible explanation of the close similarity of the $\Delta G_{gas+solv}$ values derived from the two MD simulations is that the binding pocket of *Bmor*PBP1 can accommodate multiple conformations with nearly the same binding affinity. This suggestion is supported by the MD study of Gräter *et al.* [189] on the release pathways of bombykol from *Bmor*PBP1 in which Ser9 and Trp37 were shown to play an important role. Further support comes from crystallographic work of Lautenschlager *et al.* [70] who found that *BmorPBP1* can accommodate more than one odorant molecule in its binding pocket. These results as well as those derived from the docking simulations described in Section 3.3.2 do not exclude the possibility of the presence of sub-sites in the binding pocket of *Bmor*PBP1 representing low energy minima, which may be functionally involved in ligand binding and/or release.

The two MD simulations of *Msex*GOBP2 gave a free energy of binding in favour of *Msex*GOBP2[b]. The differences in $\Delta G_{gas+solv}$ in the two simulations are attributable almost exclusively to the contribution of the electrostatic energy.

A comparison of the lowest binding free energies of the *B. mori* and *M. sexta* OBPs binding proteins with bombykal are shown in Table 3.17.

| | *Bmor*PBP [b] | | *Msex*PBP1 | | *Bmor*GOBP2 | | *Msex*GOBP2 [b] | |
|---|---|---|---|---|---|---|---|---|
| Contrib.[a] | Δ value [b] | σ [c] | Δ value [b] | σ [c] | Δ value [b] | σ [c] | Δ value [b] | σ [c] |
| $\Delta H_{vdW}$ | -41.5 | 2.6 | -42.1 | 2.7 | -43.3 | 2.0 | -42.7 | 2.2 |
| $\Delta H_{elec}$ | -1.0 | 3.4 | -4.7 | 4.0 | -4.4 | 4.3 | -6.1 | 3.9 |
| $\Delta H_{gas}$ | -42.6 | 4.2 | -46.8 | 4.9 | -47.7 | 4.9 | -48.8 | 4.2 |
| $\Delta G_{GB}$ | 13.8 | 2.8 | 17.3 | 3.5 | 17.9 | 3.3 | 16.0 | 2.7 |
| $\Delta G_{np}$ | -5.9 | 0.3 | -6.0 | 0.3 | -6.2 | 0.2 | -6.0 | 0.2 |
| $\Delta G_{solv}$ | 7.9 | 2.7 | 11.3 | 3.4 | 11.7 | 3.2 | 10.0 | 2.7 |
| $\Delta G_{gas+solv}$ | -34.6 | 2.5 | -35.5 | 2.7 | -36.0 | 2.5 | -38.8 | 2.5 |
| $\Delta TS_{total}$ | -21.5 | 3.4 | -20.8 | 3.0 | -22.4 | 1.0 | -19.3 | 2.1 |
| $\Delta G_{total}$ | -13.1 | | -14.7 | | -13.6 | | -19.5 | |

**Table 3.17** **Free energy components of OBP-bombykal complexes**
Notation with regard to superscripts [a], [b] and [c] is given in Table 3.15, p. 138

The $\Delta G_{gas+solv}$ results show overall the same trend as that observed in the case of the *B. mori* and *M. sexta* OBP-bombykol complexes and the binding affinities to bombykal are therefore in the order of:

*Msex*GOBP2 > *Bmor*GOBP2 > *Msex*PBP1 > *Bmor*PBP1

In all complexes, van der Waals interactions ($\Delta H_{vdW}$), and to a much lesser effect, the non-polar part of solvation free energy ($\Delta G_{np}$) and the electrostatic energies ($\Delta H_{elec}$) make the greatest contribution to the binding free energies. However, the $\Delta G_{gas+solv}$ values of *Bmor*GOBP2, *Msex*PBP1 and *Bmor*PBP1 do not differ

significantly and may well be within the range of experimental error. Taking into account entropic contributions the order of binding affinities becomes:

*Msex*GOBP2 > *Msex*PBP1 > *Bmor*GOBP2 > *Bmor*PBP1

Much longer MD simulations would be required to establish convergence in entropic values.

A comparison of the $\Delta G_{gas+solv}$ values of the OBPs in complex with bombykol and bombykal complexes shows that $\Delta G_{gas+solv}$ are consistently lower than the corresponding ones of the bombykol complexes as shown in Table 3.18.

| | ***Bmor*PBP1** | ***Msex*PBP1** | ***Bmor*GOBP2** | ***Msex*GOBP2** |
|---|---|---|---|---|
| $\Delta G_{gas+solv\ (BOM)}$ | -36.6 | -37.8 | -43.7 | -44.9 |
| $\Delta G_{gas+solv\ (BAL)}$ | -34.6 | -35.5 | -36.0 | -38.8 |

**Table 3.18** **$\Delta G_{gas+solv}$ of *B. mori* and *M. sexta* OBPs in complex with bombykol (BOM) and bombykal (BAL)**

The higher affinity for bombykol is thus accounted for by the more favourable contributions of the gas-phase electrostatic energy and the polar part of the solvation energy ($\Delta H_{elec}$+ $\Delta G_{GB}$) to the binding equilibrium.

### *3.4.3.2 MM-GB(PB)SA analysis and per-residue decomposition: bombykol*

The selectivity of each protein towards bombykol and bombykal was further analysed by means of per-residue decomposition of the interaction energies (Section 2.2.5.1). Per-residue decomposition calculates the energy contribution of individual residues by summing their interactions over all residues in the system. Such decomposition, which is performed on $G_{gas+solv}$ and not on $G_{total}$, can provide useful insights on type and strength of interactions between protein residues and the ligand. Per-residue decomposition of the *B. mori* and *M. sexta* proteins in complex with bombykol are shown in Table 3.19.

| *Bmor*PBP1 | | *Msex*PBP1 | | *Bmor*GOBP2 | | *Msex*GOBP2 | |
|---|---|---|---|---|---|---|---|
| Residue | $G_{gas+solv}$ | Residue | $G_{gas+solv}$ | Residue | $G_{gas+solv}$ | Residue | $G_{gas+solv}$ |
| Leu8 | -1.2 | Leu8 | | Val8 | -1.0 | Val8 | -0.9 |
| Ser9 | -0.8 | Cys9 | | Thr9 | -0.8 | Thr9 | -0.7 |
| Phe12 | -1.9 | Phe12 | -1.5 | Phe12 | -1.1 | Phe12 | -1.1 |
| Phe36 | | Phe36 | -0.8 | Phe36 | | Phe36 | |
| Trp37 | | Trp37 | -0.9 | Trp37 | | Trp37 | |
| Ile52 | -1.0 | Ile52 | -0.9 | Ile52 | -1.1 | Ile52 | -1.4 |
| Ser56 | -1.1 | Ser56 | | Ser56 | | Ser56 | |
| Met61 | -0.8 | Met61 | | Leu61 | | Leu61 | |
| Leu68 | | Leu68 | -1.4 | Met68 | -1.3 | Ile68 | -1.3 |
| Leu90 | | Leu90 | -1.2 | Met90 | | Met90 | |
| Ile91 | | Leu91 | -0.7 | Val91 | | Val91 | |
| Val94 | -1.2 | Val94 | | Ile94 | -1.3 | Ile94 | -1.5 |
| Glu98 | | Glu98 | | Glu98 | -0.8 | Glu98 | -2.1 |
| Thr110 | | Lys110 | | Arg110 | -2.5 | Arg110 | -2.2 |
| Thr111 | -1.0 | Thr111 | | Val111 | -1.2 | Val111 | -1.2 |
| Val114 | -1.0 | Ile114 | | Val114 | -1.3 | Ile114 | -1.3 |
| Ala115 | | Ala115 | -0.8 | Ala115 | | Ala115 | -0.8 |
| Phe118 | -1.2 | Phe118 | -1.1 | Phe118 | -0.9 | Phe118 | -1.2 |

**Table 3.19** **Per-residue decomposision of $G_{gas+solv}$ of *B. mori* & *M. sexta* OBP-bombykol complexes**

The table summarises the residues making the largest contributions to the free energies of binding. Values higher than a cutoff of -0.7 kcal $mol^{-1}$ are not shown. The calculated values include contributions from van der Waals energies, electrostatic interactions and the polar and non-polar contributions to solvation free energy. The complete data can be found in Supplementary Materials: Per-Residue Decomposition. For almost all residues shown in the table, van der Waals interactions are dominant. Exceptions are the residues that have been shown to

be involved in H-bond formation in which cases Coulombic energies are the dominant contributors to $G_{gas+solv}$.

A graphical representation of the data is shown in Figure 3.21, which depicts the difference:

$$\Delta G^{residue}_{gas+solv} = (G^{residue}_{gas+sol})_{BmorOBP} - (G^{residue}_{gas+solv})_{MsexOBP}$$

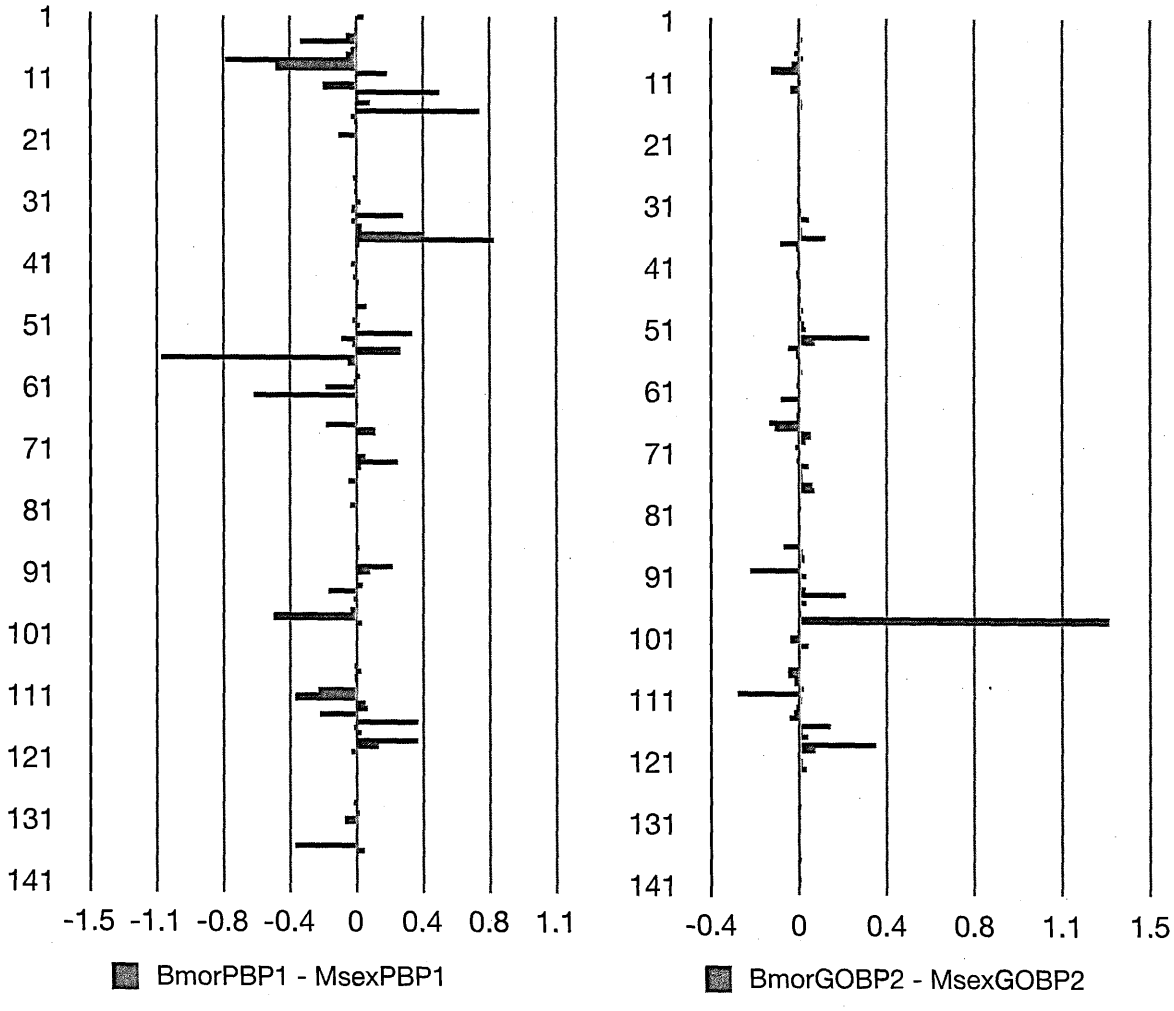



**Figure 3.21** **Per residue decomposition of binding free energies of moth OBPs with bombykol**

Comparison between (a) the PBP - bombykol complexes & (b) the GOBP-bombykol complexes

Comparison of the interactions between *Bmor*PBP1 and *Msex*PBP1 with bombykol shows that in the course of the 10ns MD trajectories, Ser56 and to a lesser extent Leu8, Ser9, Phe12, Met61, Lys98 and Val135 favour *Bmor*PBP1. For

*Msex*PBP1 the most favourable interaction, in comparison to *Bmor*PBP1, is predominantly with Trp37, and to some measure with Leu68, Ala73, Met74 and Leu90. Ser56 and Trp37 are involved in H-bonding as seen earlier (Section 3.4.2.2). All of the residues mentioned above are conserved in the two proteins with the exception of Ser9 of *Bmor*PBP1 which is replaced by Cys9 in *Msex*PBP1. As expected, the residues involved in H-bond formation make the largest contribution to the "effective energy".

Likewise, in the case of *Bmor*GOBP2 and *Msex*GOBP2, with the exception of Glu98 which in the case of *Msex*GOBP2 makes a much greater contribution towards the "effective energy", all other residues of the two proteins seem to contribute equally to the binding of bombykol.

#### *3.4.3.3 MM-GB(PB)SA analysis and per-residue decomposition: bombykal*

The data derived from a similar treatment of the 10ns MD trajectories involving the *B. mori* and *M. sexta* pheromone and general odorant binding proteins are summarised in Table 3.20. No significant electrostatic contributions to the per-residue $G_{gas+solv}$ were discernible and interactions with bombykal were predominantly hydrophilic.

Per-residue $\Delta G_{gas+solv}$ values reflecting the differences of the relative contribution of each residue to the free energy of binding between bombykal and each of the pairs of PBPs (*Bmor*PBP1, *Msex*PBP1) and GOBPs (*Bmor*GOBP2, *Msex*GOBP2) are depicted graphically in Figure 3.23.

| *Bmor*PBP1 | | *Msex*PBP1 | | *Bmor*GOBP2 | | *Msex*GOBP2 | |
|---|---|---|---|---|---|---|---|
| Residue | $G_{gas+solv}$ | | $G_{gas+solv}$ | | $G_{gas+solv}$ | | $G_{gas+solv}$ |
| Leu8 | -1.0 | Leu8 | | Val8 | | Val8 | -0.9 |
| Ser9 | | Cys9 | | Thr9 | | Thr9 | -0.8 |
| Phe12 | -1.8 | Phe12 | | Phe12 | -1.2 | Phe12 | -1.1 |
| Phe36 | | Phe36 | | Phe36 | | Phe36 | -0.7 |
| Ile52 | -1.0 | Ile52 | -0.9 | Ile52 | -1.4 | Ile52 | -1.4 |
| Met61 | -0.9 | Met61 | -1.1 | Leu61 | | Leu61 | |
| Leu62 | | Ile62 | -1.1 | Met62 | | Leu62 | |
| Asn67 | | Lys67 | | Arg67 | -0.8 | Arg67 | |
| Leu68 | -1.0 | Leu68 | -1.5 | Met68 | -1.4 | Ile68 | -1.9 |
| Leu90 | | Leu90 | | Met90 | -0.8 | Met90 | |
| Val94 | -1.3 | Val94 | -1.1 | Ile94 | -1.3 | Ile94 | -1.3 |
| Trp110 | | Lys110 | | Arg110 | -2.3 | Arg110 | |
| Thr111 | -0.7 | Thr111 | -1.2 | Val111 | -1.1 | Val111 | -1.6 |
| Val114 | -1.0 | I114 | -1.8 | Val114 | -1.0 | Val114 | -1.1 |
| Ala115 | -0.7 | Ala115 | | Ala115 | -0.7 | Ala115 | |
| Phe118 | -1.2 | Phe118 | -0.9 | Phe118 | -1.0 | Phe118 | -1.1 |

**Table 3.20** **Per-residue decomposision of $G_{gas+solv}$ of *B. mori* & *M. sexta* OBP-bombykal complexes**

$G_{gas+solv}$ in kcal mol$^{-1}$; $G_{gas+solv}$ cutoff -0.7 kcal mol$^{-1}$

The data shown in Table 3.20 and Figure 3.22 show that Ser9 makes a greater contribution to the free energy of binding in *Bmor*PBP1 than in *Msex*PBP1. Trp37 likewise but to a much lesser extent. The comparison of the two GOBPs shows that the residues shown in Table 3.19 make nearly equal contributions to binding with the exception of Arg67 and Arg110 that make much larger contributions in favour of *Msex*GOBP2 although the latter much more so.

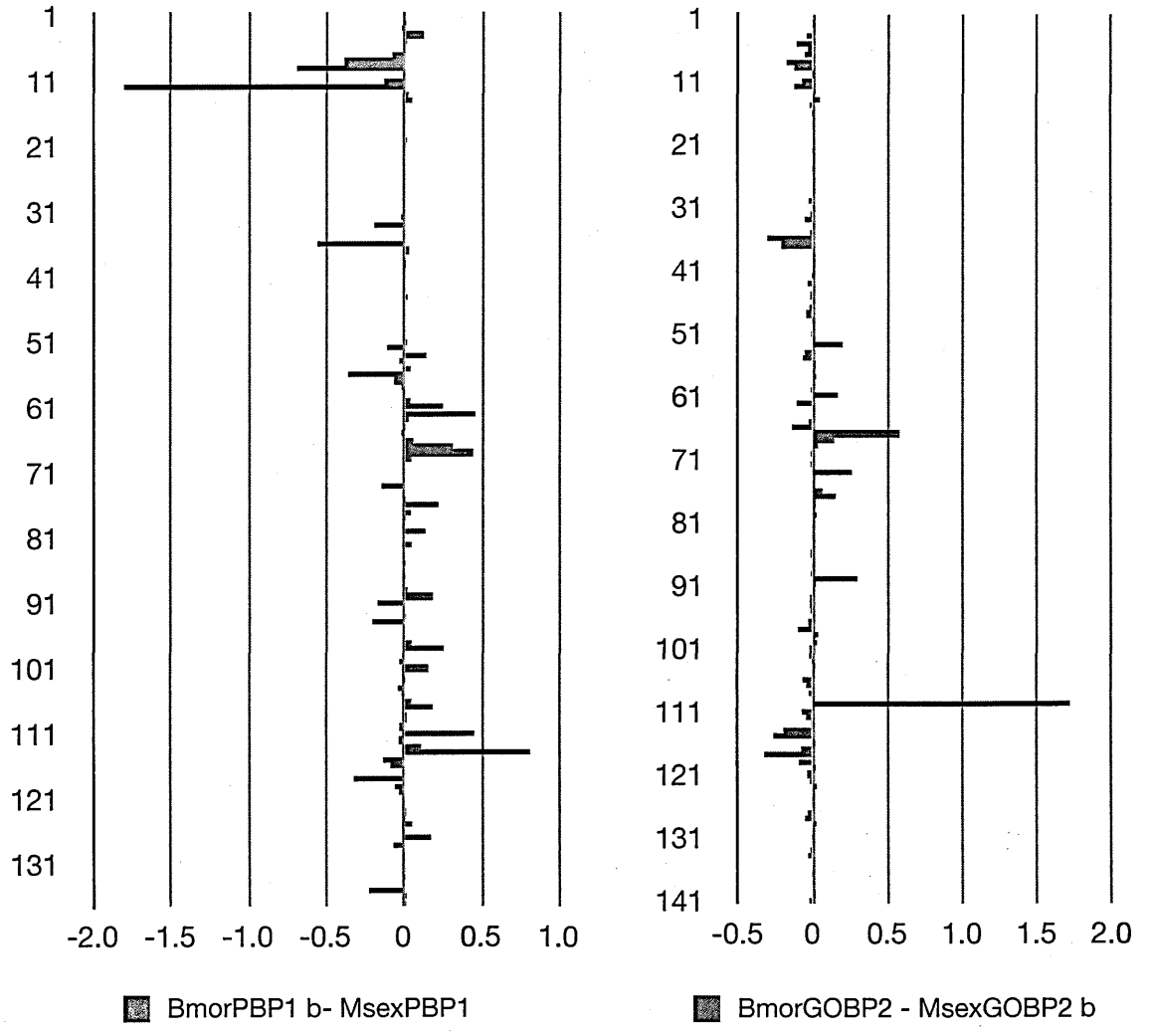



**Figure 3.22** **Per residue decomposition of binding free energies of moth OBPs with bombykal**

Comparison between (a) the PBP - bombykol complexes & (b) the GOBP-bombykol complexes

A similar comparison between the PBPs and GOBPs of the two species reveals preferential binding of bombykol to the GOBPs. Of the residues that make the largest contribution to the free energies of binding with bombykol, Phe12, Trp37, Ser 56, Glu98, Ala115 and Phe118 are shared by all four proteins.

A comparison of Tables 3.19 and 3.20 helps to identify those residues of all four OBPs whose contributions to the binding free energies of bombykol and bombykal is the largest. These are shown in Table 3.21 below.

| | *Bmor*PBP1 | | *Msex*PBP1 | | *Bmor*GOBP2 | | *Msex*GOBP2 | |
|---|---|---|---|---|---|---|---|---|
| | Residue | $G_{gas+solv}$ | Residue | $G_{gas+solv}$ | Residue | $G_{gas+solv}$ | Residue | $G_{gas+solv}$ |
| BOM | Phe12 | -1.9 | Phe12 | -1.5 | Phe12 | -1.1 | Phe12 | -1.1 |
| BAL | Phe12 | -1.8 | Phe12 | | Phe12 | -1.2 | Phe12 | -1.1 |
| BOM | Ile52 | -1.0 | Ile52 | -0.9 | Ile52 | -1.1 | Ile52 | -1.4 |
| BAL | Ile52 | -1.0 | Ile52 | -0.9 | Ile52 | -1.4 | Ile52 | -1.4 |
| BOM | Leu68 | | Leu68 | -1.4 | Met68 | -1.3 | Ile68 | -1.3 |
| BAL | Leu68 | -1.0 | Leu68 | -1.5 | Met68 | -1.4 | Ile68 | -1.9 |
| BOM | Val94 | -1.2 | Val94 | | Ile94 | -1.3 | Ile94 | -1.5 |
| BAL | Val94 | -1.3 | Val94 | -1.1 | Ile94 | -1.3 | Ile94 | -1.3 |
| BOM | Thr111 | -1.0 | Thr111 | | Val111 | -1.2 | Val111 | -1.2 |
| BAL | Thr111 | -0.7 | Thr111 | -1.2 | Val111 | -1.1 | Val111 | -1.6 |
| BOM | Val114 | -1.0 | Ile114 | | Val114 | -1.3 | Ile114 | -1.3 |
| BAL | Val114 | -1.0 | I114 | -1.8 | Val114 | -1.0 | Val114 | -1.1 |
| BOM | Phe118 | -1.2 | Phe118 | -1.1 | Phe118 | -0.9 | Phe118 | -1.2 |
| BAL | Phe118 | -1.2 | Phe118 | -0.9 | Phe118 | -1.0 | Phe118 | -1.1 |

**Table 3.21** **Residues making the largest contribution to binding free energies**
$G_{gas+solv}$ in kcal mol$^{-1}$; $G_{gas+solv}$ cutoff -0.7 kcal mol$^{-1}$

All of these residues, with the exception of Thr111, are non-polar and are involved in hydrophobic and van der Waals interactions with the ligands. The observed differences in the free energies of binding of these proteins thus arise from the relative strengths of these interactions. Specificity to bombykol and bombykal is conferred from hydrogen bonding interactions between the polar moieties of these ligands and residues Ser56, Glu98 and Arg110 of the proteins.

On the basis of the total free energies of binding as well as the corresponding data concerning the per-residue decomposition of $G_{gas+solv}$, the GOBPs seem to discriminate strongly bombykol to bombykal. The PBPs, on the other hand, are predicted to have a discriminatory capacity for bombykol but much less so for bombykal. If entropic contributions to the total binding energy are taken into

account, the two *M. sexta* OBPs would be predicted to have a higher binding affinity towards bombykal than the corresponding *B. mori* proteins.

To determine whether steric factors play a role in restricting the movement of the ligands within the binding cavities, the volumes of the binding pockets of the proteins were measured using the CASTp program (Section 2.2.7). The results are summarised in Table 3.22.

| Protein | PDB:ID | N_mth | Area_sa (Å$^2$) | Area_ms (Å$^2$) | Vol_sa (Å$^3$) | Vol_ms (Å$^3$) | Length (Å) |
|---|---|---|---|---|---|---|---|
| *Bmor*PBP1 | 1dqe:A | 3 | 341 | 665 | 203 | 889 | 339 |
| *Msex*PBP1 | NA | 2 | 368 | 762 | 218 | 1004 | 380 |
| *Bmor*GOBP2 | 2wc6:A | 0 | 236 | 504 | 141 | 636 | 238 |
| *Msex*GOBP2 | NA | 0 | 272 | 544 | 190 | 752 | 272 |

**Table 3.22** **Binding pocket dimensions of *B. mori* and *M. sexta* OBPs**
N_mth: number of openings to the external molecular surface;
Area(Vol)_sa: pocket solvent-accessible area (volume);
Area(Vol)_ms: pocket molecular surface area (volume)

The data show that the binding cavities of *Bmor*PBP1 and *Msex*PBP1 are open to bulk solvent, contrary to the binding sites of *Bmor*GOBP2 and *Msex*GOBP2. In addition, the volume and lengths of the cavities circumscribing the binding sites of *Bmor*PBP1 and *Msex*PBP1 are ~30% greater than those of *Bmor*GOBP2 and *Msex*GOBP2, respectively. It is possible that the smaller binding cavities of *Bmor*GOBP2 and *Msex*GOBP2 may hinder the rotational movement of the ligand due to steric clashes with the side chains of the binding site residues.

### 3.5.1 Summary of findings

Docking simulations which *B. mori* and *M. sexta* OBPs showed that ligand binding is quite promiscuous with ligands able to attain several poses within the binding pockets of the proteins. This is consistent with the results reported in the literature

[98]. Both docking and MD simulations, as well as the results obtained by calculating the molecular interaction properties of the binding sites of the proteins (Section 3.3.3), provided evidence of the presence of two distinct low energy "hot spots" within the binding sites of the proteins, one proximal to residues 56, 98, 110 and the other proximal to residues 36, 37 and 9 (Table 3.23 below). Docking and MD simulations, showed that, at these "hot spots", binding of ligands with polar head groups is equally favourable. Residues involved in hydrogen bonding include Ser9 or Thr9 at one "hot spot" and Ser56 or Glu98 at the other. Hydrogen bond formation was also shown to involve Arg110 of *Bmor*GOBP2 and *Msex*GOBP2. Transient hydrogen bonding with other residues was also observed.

| **Protein** | **Residue** | | | | | |
|---|---|---|---|---|---|---|
| *Bmor*PBP1 | Ser9 | Phe36 | Trp37 | Ser56 | Glu98 | Trp110 |
| *Msex*PBP1 | Cys9 | Phe36 | Trp37 | Ser56 | Glu98 | Lys110 |
| *BmorGOBP2* | Thr9 | Phe36 | Trp37 | Ser56 | Glu98 | Arg110 |
| *MsexGOBP2* | Thr9 | Phe36 | Trp37 | Ser56 | Glu98 | Arg110 |

**Table 3.23 Residues involved in hydrogen-bonding and/or stabilising interactions**

Both MD and docking simulations showed that binding affinities are dominated by hydrophobic interactions suggesting that ligand selectivity may depend on a very fine interplay of hydrogen bonding, electrostatic and weak interactions such as π-π interactions, etc. In particular, Phe36 and Trp37 appear to play an important role in stabilising protein-ligand interactions. The size of the ligand and the relative position of unsaturated double bonds and/or aromatic moieties was also found to play a role in determining the orientation and H-bonding capacity of ligands.

Consistent with the crystallographic work of Lautenschlager *et al.* [70] it was found that *BmorPBP1* can accommodate more than one ligand of small size in its binding pocket but would not tolerate a second ligand molecule when in complex

with its physiological pheromone. It is possible that in view of the high concentration of PBPs in the sensillum lymph and of the fact that only a fraction of these proteins is needed to transport the pheromone to the sensory neurons, the vast majority of the PBP molecules may be involved in the sequestration of other odorant molecules.

The promiscuity in ligand binding shown by these OBPs coupled with some selectivity for molecules with specific geometries could well be an evolutionary adaptation supporting the dual "carrier-scavenger" model of OBP function.

A comparison of the ligand binding affinities for the *B. mori* and *M. sexta* OBPs obtained from docking simulations shows that the *B. mori* OBPs bind alcohols with greater affinity than the corresponding *M. sexta* OBPs. *Msex*PBP1 was shown to bind aldehyde ligands with almost equal or better affinity than the *B. mori* OBPs and with considerably more affinity than *Msex*GOBP2. In general, the lowest energy minima reached for the GOBP-ligand complexes were consistently lower than the corresponding PBP ones. However, the selectivity of the *B. mori* and *M. sexta* OBPs towards the tested ligands cannot be determined conclusively from the free energies of binding obtained from docking simulations as they are within the standard deviation of the AutoDock force field. Exceptionally, retinol was found to bind more strongly than any other ligand and discriminate between the *B. mori* and *M. sexta* OBPs (see Table 3.3 and Figure 3.7).

The molecular interaction energies of the binding cavities of the orthologous pairs of *B. mori* and *M. sexta* OBPs showed distinct regions in which binding is more favourable in one ortholog compared to the other. Such differences in molecular interaction energies can be exploited for fragment-based design of ligands with a potential to disrupt olfactory function.

SMD simulations identified one ligand entry-exit pathway in *Bmor*GOBP2, contrary to the two identified in *Bmor*PBP1. It has been suggested that multiple gateways not only lower the energy barrier to ligand entry/exit but also ensure the robustness of the binding properties of *Bmor*PBP1 against point mutations that could eliminate one of the routes [189]. Multiple longer SMD simulations would be necessary to exclude the possibility of more than one entry/exit gateways.

Binding affinities calculated from MD simulations using the MM-GBSA model predicted "effective" free energies of binding of the moth OBP-bombykol complexes in the order of:

*Msex*GOBP2 > *Bmor*GOBP2 > *Msex*PBP1 > *Bmor*PBP1

Similarly, the "effective" free energies of binding of the moth OBP-bombykal complexes were in the order of:

*Msex*GOBP2 > *Bmor*GOBP2 > *Msex*PBP1 > *Bmor*PBP1

The results shown in Table 3.17 indicate the GOBPs have considerably higher affinity towards bombykol, and to a lesser extent bombykal, than the corresponding PBPs. The crystal structures of *Bmor*PBP1 and *Bmor*GOBP2 in complex with bombykol have buried water molecules involved in bridged H-bonds between the hydroxyl group of bombykol and Ser56 of *Bmor*PBP1 and Glu98, Arg110 of *Bmor*GOBP2, which suggests that water may play an important role in binding affinity as well as binding selectivity. Longer MD simulations, as well as simulations involving the buried water molecules would be needed to obtain better estimates of the binding affinities of these proteins.

*Chapter 4*

# *Deciphering the specificity of mosquito OBPs*

# 4.1 Mosquito OBPs

## 4.1.1 Background

Despite the increasing number of mosquito OBPs that have been identified in recent years only few of them have been structurally characterised (Table 1.2). *Agam*OBP4 is a member of the medium length classical OBPs having structural similarity with LUSH with which it shares 41% of its residues [194]. *Agam*OBP7 represents a new structural class of OBPs which is characterised by 8 cysteines and four disulfide bridges [55]. *Agam*OBP47 is a member of the C-plus class of OBPs having a longer sequence than the classical OBPs and containing six disulfide bridges [56]. *Agam*OBP22 although it has the characteristic three disulfide bridges of the classical OBPs its structural similarity with them is little [11]. The sequence similarity of these OBPs with *Agam*OBP1 is less than 21%. The latter is a member of the medium size classical OBPs having very high sequence similarity (> 80% identity) with *A. agypti* and *C. quinquefasciatus* orthologs and almost identical structures as shown in Figure 4.1, despite the very different chemical ecologies of these three species. There is experimental evidence showing that *Agam*OBP1 plays a role in the olfactory response to indole which is a human volatile and an oviposition pheromone. Likewise, the elimination of *Cqui*OBP1 was shown to result in reduced electrophysiological response to oviposition pheromone and other odorants that bind to this protein. *Agam*OPB1, *Aaeg*OBP1 and *Cqui*OBP1 are the subject matter of this study.

### 4.1.1.1 Structural features of the mosquito OBP orthologs

The structural features of *Agam*OBP1, *Aaeg*OBP1 and *Cqui*OBP1 have been analysed extensively [67], [68], [69]. All three proteins have been crystallised as

[11] No publication was associated with this crystal structure at the time of writing of this thesis

dimers suggesting that dimerisation may play a physiologically relevant role. The physiological relevance of the dimerisation is further explored in Section 4.3.3.

Of the above three models, only *Cqui*OBP1 has been crystallised bound to a biologically relevant ligand, namely the oviposition pheromone (*5R,6S*)-6-acetoxy-5-hexadecanolide (MOP) and *Agam*OBP1 with the insect repellent *N,N*-Diethyl-*m*-toluamide (DEET). *Agam*OBP1 and *Aaeg*OBP1 have also been crystallised bound to a PEG molecule from the crystallisation solution.


**Figure 4.1** **Structural alignment of *Agam*OBP1, *Aaeg*OBP1 and *Cqui*OBP1**

In these OBPs, helices 4 and 5 of each monomeric subunit form a hydrophobic tunnel. The residues lining the tunnel of subunit A of these proteins are Leu75, Leu78, His79 and Leu82 from helix 4, Met93, Gly94 and Leu98 from helix 5, as

well as Trp116 from helix 6[12]. The two tunnels meet to form the dimer interface, which interconnects the central binding cavities of the two subunits. In the *Agam*OBP1- and *Aaeg*OBP1-PEG complexes, a single 55-80 atom molecule of PEG runs through the interconnected cavities of the two sub-units and exits through openings formed by helices 1, 3 and 4. Contrary to this, in the case of *Cqui*OBP1-MOP complex, two molecules of MOP are bound, one in each monomeric sub-unit. The lactone/acetyl ester "head" of the ligand occupies approximately half of the central cavity of the receptor, whereas its long lipid "tail" is buried in the tunnel formed by helices 4 and 5. The crystal structure of 3ogn also reveals that the electron density around the lactone ring is not very well defined suggesting that this part of the ligand can attain different conformations inside the cavity. The binding site thus appears not to impose strict conformational constraints on the ligand. This relative plasticity of the OBP binding site is not a distinctive feature of *Cqui*OBP1 alone and has been observed in the case of *Bmor*PBP1 as well [70]. Despite this, MOP makes numerous hydrophobic and van der Waals interactions inside the binding cavity. The acetyl ester moiety of the ligand makes hydrophobic interactions with Tyr12, Leu82, Ala90, Met93, His123 and Phe125, while the lactone ring interacts with Leu17, Leu21, Leu82, His113, Tyr124 and Phe125. These residues are also conserved in *Agam*OBP1 and *Aaeg*OBP1 with the exception of Leu15 and Leu19 which are replaced by the bulkier but also hydrophobic Phe17 and Met21 in the case of *Aaeg*OBP1.

In the *Agam*OBP1-DEET complex, two ligand molecules occupy the tunnels interconnecting the dimeric subunits as shown in Figure 4.2.

---

[12] Henceforth, unless otherwise stated, the quoted residue numbers correspond to the structural alignment of Figure 4.1

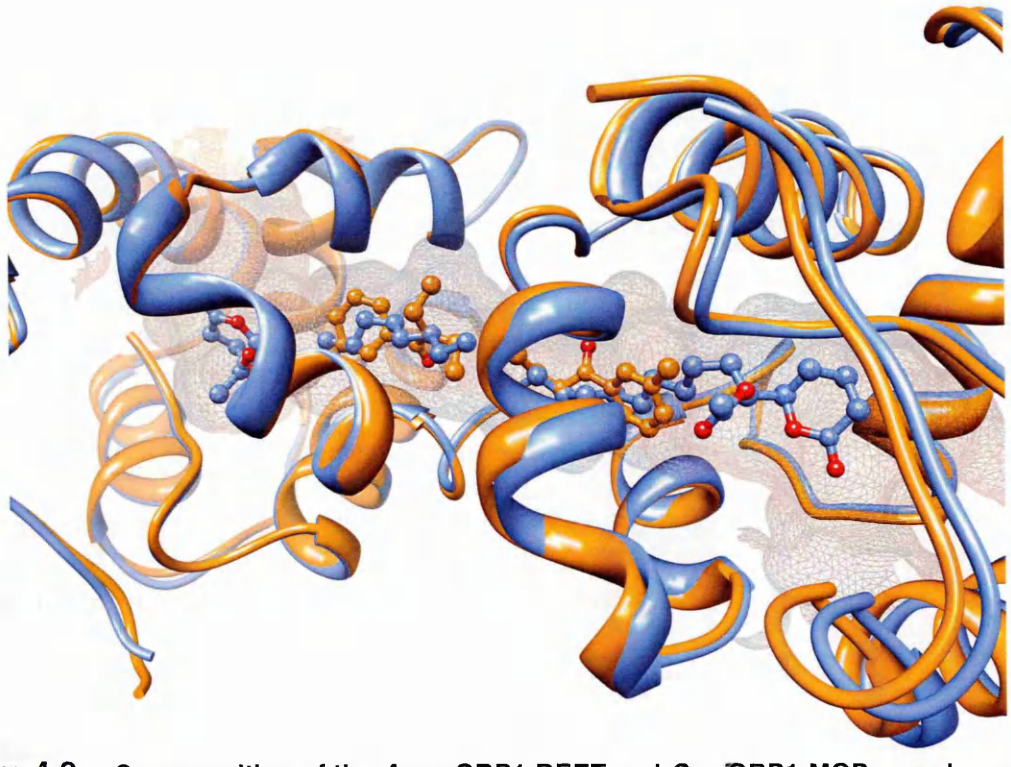


**Figure 4.2** **Superposition of the *Agam*OBP1-DEET and *Cqui*OBP1-MOP complexes**
*Cqui*OBP1-MOB (blue); *Agam*OBP1-DEET (orange)

DEET makes numerous van der Waals and hydrophobic interactions with residues of helix α4 (Leu73, Leu76, His77, Leu80), helices α5 (Ala88, Met89, Met91, Gly92) and α6 (Trp114), as well as residues Leu96, Lys93, Arg94 and Leu96 belonging to the second subunit (chain B). It also forms a hydrogen bond from its amide oxygen to a water molecule, which in turn interacts with the carbonylic oxygen of Gly92 and with the NE1 nitrogen atom of Trp114, alternately [66]. The figure above is a superimposition of the *Agam*OBP1-DEET and *Cqui*OBP1-MOP complexes and depicts the central cavities and the tunnels interconnecting the two protein subunits as well as the relative positions of the bound ligands. It is clearly shown that DEET lies on either side of the tunnel interconnecting the two subunits with its carbonyl oxygen directed towards the interface, whereas the bulky lactone moiety of MOP lies well within the main binding cavities of the monomeric subunits and its hydrophobic tail extends towards the dimeric interface.

Another particular structural feature merits attention. The C-terminus of the three OBPs is locked onto helices 1 and 3 through interactions of the terminal carboxylate with the hydroxyl group of Tyr56 and the δ-N of His25 forming a "lid" on the central cavity. In the case of *Cqui*OBP1, this lid makes extensive contacts with both the lactone and acetyl ester groups of MOP and may thus play a functional role in constraining the ligand within the protein cavity. It has been suggested that the "lid" is pH sensitive and a drop in pH may result in the disruption of the hydrogen bond network that holds it in place resulting in the release of the ligand from the complex.

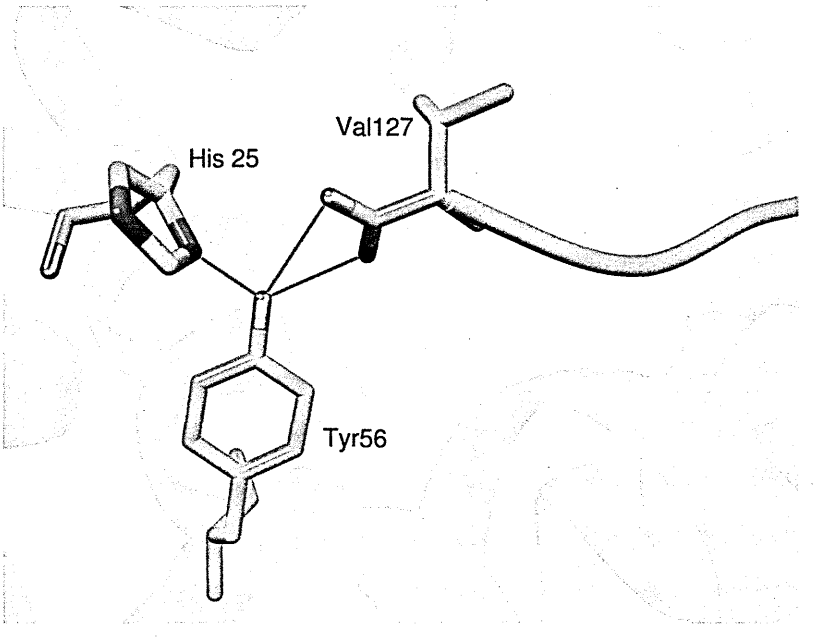



**Figure 4.3** **H-bond network locking the terminal "lid" of *Cqui*OBP1 in position**

# 4.2 Docking simulations

## 4.2.1 Methods

The procedures followed for the docking of ligand libraries and cross docking were the same as those described in Sections 2.2.2.1 and 2.2.2.2.

#### 4.2.1.1 OBP models used in docking simulations

Docking simulations were conducted with the known crystal structures of *A. gambiae* (2erb), *A. aegypti* (3k1e) and *C. quinquefasciatus* (3ogn). Table 4.1 shows the very high sequence identity of these proteins.

| OBP | Model ID | 2erb | 3k1e | 3ogn |
|---|---|---|---|---|
| *Agam*OBP1 | 2erb | 100 | | |
| *Aaeg*OBP1 | 3k1e | 81.6 | 100 | |
| *Cqui*OBP1 | 3ogn | 90.3 | 87.1 | 100 |

**Table 4.1 Sequence identity of mosquito OBPs used in docking**

#### 4.2.1.2 Pheromones and other ligands used in virtual and cross docking

The ligands comprising the docking library represent the odour space of *D. melanogaster* and *A. gambiae* [195] [196]. The total number of ligands used for screening was 355 (Supplementary Material: Mosquito OBPs: List of compounds contained in the docking library). They were grouped according to structural and/or functional group characteristics as it has been found that some classes of odorants elicit stronger responses than others. Fit quality scores were derived as described in Section 3.3.2.

## 4.3 Docking results

### 4.3.1 Screening of ligand libraries

The ligands were grouped according to functional groups and the detailed results of virtual screening are presented in (Supplementary Material: Mosquito OBPs: Results of virtual screening). There was no discernible differentiation of binding affinities for cyclic, aromatic and S-containing compounds, alcohols, aldehydes, ketones, acids and esters. Figure 4.4 shows free energies of binding and the FQ scores of ligands with FQ score > 6.5:

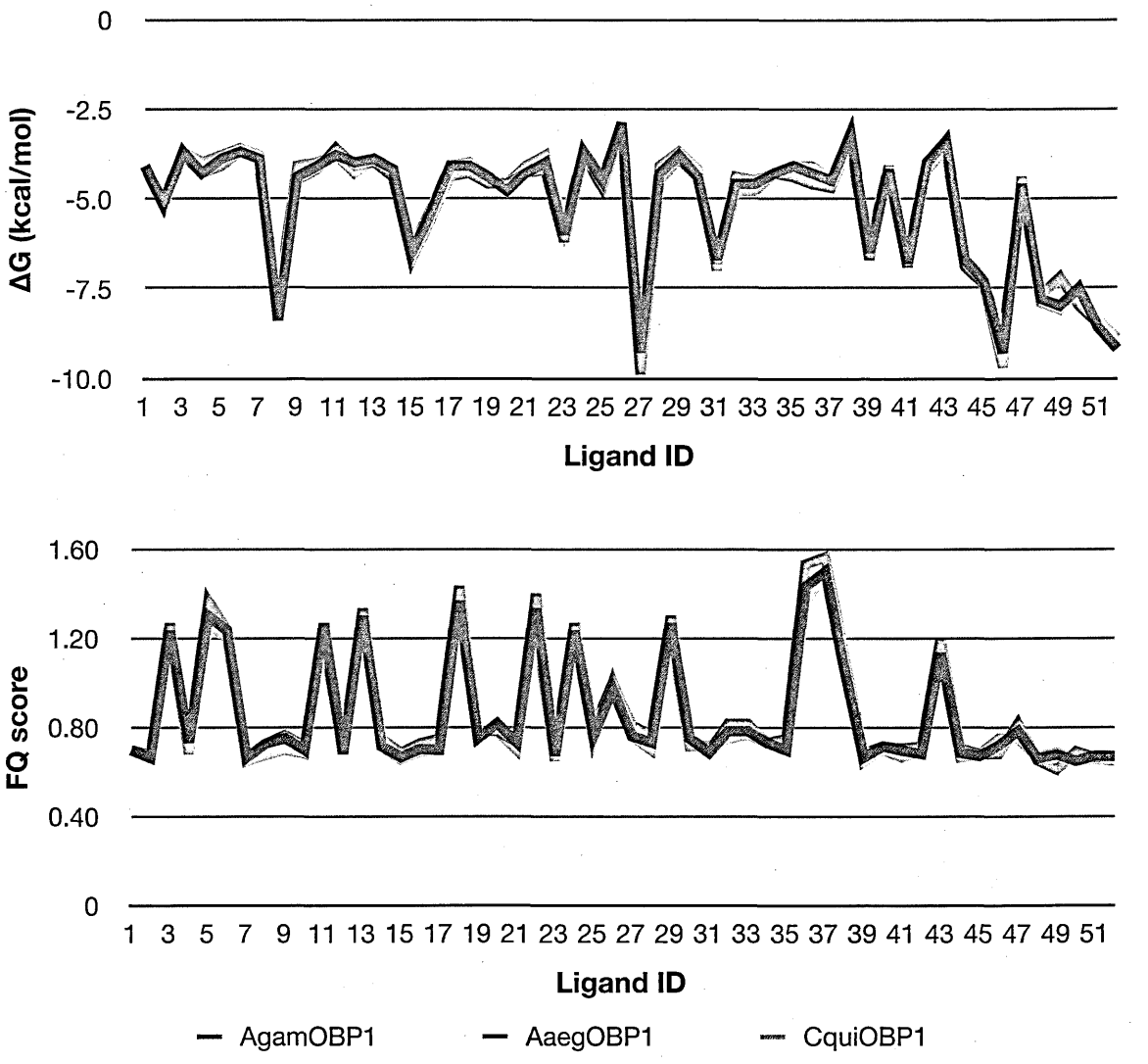



**Figure 4.4** **Free energies of binding and FQ scores of docked ligands**

Plots of the FQ scores against the number of heavy atoms of the ligands show that optimal FQ scores are obtained for compounds with 6 heavy atoms for *Aaeg*OBP1 and *Cqui*OBP1. These ligands are identical for both OBPs. Optimal FQ scores for *Agam*OBP1 are obtained for ligands with 6 and 7 heavy atoms. The results obtained from the screening of the ligand library could be used for ligand optimisation in structure-based modelling of mosquito repellents. This is beyond the scope of this work which is mainly intended to determine specificity of binding of the OBPs under consideration. Detailed results are given in (Supplementary Material: Mosquito OBPs: Results of virtual screening).

### 4.3.2 Cross docking

The ligands used in the docking simulations were chosen amongst the ones that showed the highest binding affinities in the virtual docking simulations. An additional criterion for the choice of ligands was that they represent different structural classes. The ligands used belong to the odour space of *A. gambiae* and *D. melanogaster* with the exception of picaridine, DEET and retinol. Of these, picaridine and DEET have been identified as the best mosquito repellents and retinol has been shown to bind strongly to *A. aegypti* OBP22 in competitive binding experiments [197]. The free energies of binding are shown in Figure 4.5.

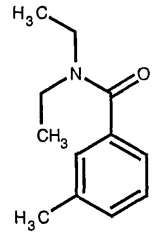
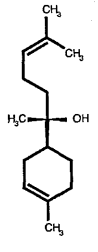
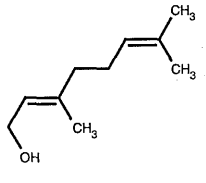
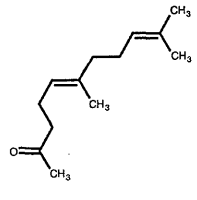
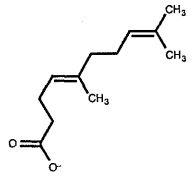
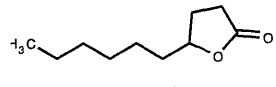
| | | | |
|---|---|---|---|
|  N,N-diethyl-3-methylbenzamide | **Emin** (kcal mol$^{-1}$)<br>[1] -6.6<br>[2] -6.7<br>[3] -6.6<br><br>**H-bonds**<br>[1] None<br>[2] None<br>[3] None |  2-(cyclohex-3-en-1-yl)-6-methylhept-5-en-2-ol | **Emin** (kcal mol$^{-1}$)<br>[1] -7.7<br>[2] -8.4<br>[3] -7.6<br><br>**H-bonds**<br>[1] Phe123<br>[2] Tyr121; Phe122<br>[3] Tyr121; Phe122 |
|  (2E)-3,7-dimethylocta-2,6-dien-1-ol | **Emin** (kcal mol$^{-1}$)<br>[1] -5.4<br>[2] -5.7<br>[3] -5.4<br><br>**H-bonds**<br>[1] Phe123<br>[2] None<br>[3] None |  (5E)-6,10-dimethylundeca-5,9-dien-2-one | **Emin** (kcal mol$^{-1}$)<br>[1] -6.7<br>[2] -7.0<br>[3] -6.55<br><br>**H-bonds**<br>[1] Trp114<br>[2] None<br>[3] Trp113 |
|  (4E)-5,9-dimethyldeca-4,8-dienoic acid | **Emin** (kcal mol$^{-1}$)<br>[1] -6.2<br>[2] -6.4<br>[3] -6.1<br><br>**H-bonds**<br>[1] None<br>[2] Phe122 (Trp113)<br>[3] None (Phe122) |  5-hexyloxolan-2-one | **Emin** (kcal mol$^{-1}$)<br>[1] -5.6<br>[2] -5.8<br>[3] -5.5<br><br>**H-bonds**<br>[1] Trp114 (polar)<br>[2] Trp113 (polar)<br>[3] None |

**Figure 4.5** **H-bond formation and binding energies of the lowest energy clusters (contd.)**

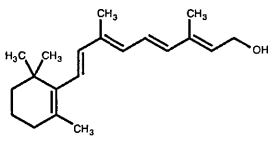
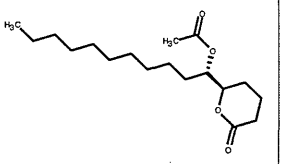
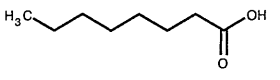
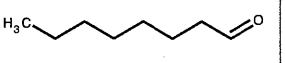
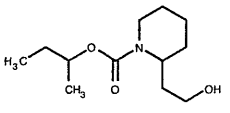
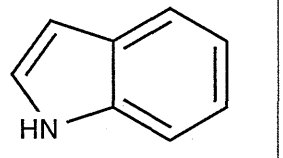
| | | | |
|---|---|---|---|
|  | **Emin** (kcal mol$^{-1}$)<br>[1] -9.56<br>[2] -10.66<br>[3] -9.66 |  | **Emin** (kcal mol$^{-1}$)<br>[1] -8.8<br>[2] -9.4<br>[3] -8.6 |
| Retinol* | **H-bonds**<br>[1] None<br>[2] None<br>[3] None | (5R,6S)-6-acetoxy-5-hexadecanolide | **H-bonds**<br>[1] None<br>[2] Phe122<br>[3] None |
|  | **Emin** (kcal mol$^{-1}$)<br>[1] -4.6<br>[2] -4.8<br>[3] -4.8 |  | **Emin** (kcal mol$^{-1}$)<br>[1] -4.18<br>[2] -4.38<br>[3] -4.1 |
| n-octanoic acid | **H-bonds**<br>[1] Trp114<br>[2] Phe122<br>[3] Trp113 | n-octanal | **H-bonds**<br>[1] None<br>[2] None<br>[3] None |
|  | **Emin** (kcal mol$^{-1}$)<br>[1] -6.5<br>[2] -6.9<br>[3] -6.5 |  | **Emin** (kcal mol$^{-1}$)<br>[1] -5.5<br>[2] -5.1<br>[3] -5.5 |
| butan-2-yl 2-(2-hydroxyethyl)piperidine-1-carboxylate | **H-bonds**<br>[1] None<br>[2] None<br>[3] Met90 | indole | **H-bonds**<br>[1] Asp70<br>[2] Leu72<br>[3] Asp69 |

**Figure 4.5 (contd.)** **H-bond formation and binding energies of the lowest energy clusters**

[1] *Agam*OBP1; [2] *Aaeg*OBP1; [3] *Cqui*OBP1

(*)(2E,4E,6E,8E)-3,7-dimethyl-9-(2,6,6-trimethylcyclohex-1-en-1-yl) nona-2,4,6,8-tetraen-1-ol

There are not discernible differences in the binding affinities of the three OBPs as the free energies of binding are within the standard deviation of AutoDock.

The docked complexes of DEET and MOP are in good agreement with the experimental structures as shown in Figure 4.6. However, in both cases the docked ligand structures do not come from the lowest energy clusters. No water molecules were included in the docking simulations and this may have an effect on the docking results. Both X-ray models (*Agam*OBP1-DEET and *Cqui*OBP1-MOP)

show the presence of structured water molecules in the binding cavities of the OBPs.

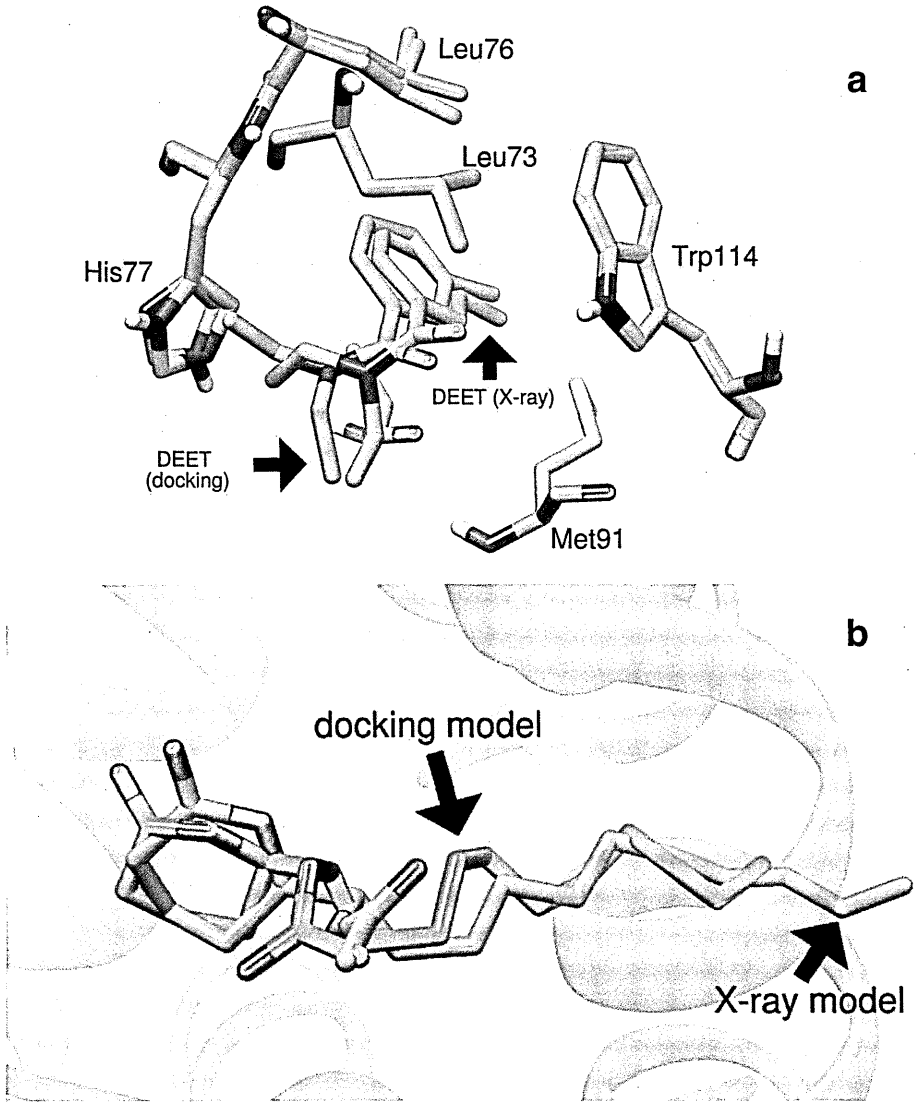



**Figure 4.6** **Superposition of docked ligands over the corresponding X-ray models**
a. *Agam*OBP1-DEET; b. *Cqui*OBP1-MOP

Interestingly, several ligands shown in Figure 4.5 formed H-bonds with the peptidic oxygen of Phe123 of *Agam*OBP1 and Phe122 of *Aaeg*OBP1 and *Cqui*OBP1. A representative structure of the *Agam*OBP1-bisabolol complex is shown in Figure 4.7. The figures show, that contrary to the *Agam*OBP-DEET complex in which the polar moiety of DEET was shown to be directed towards Trp114, the hydroxyl group of bisabolol points away from it towards the C-terminus of the protein. The

ligand is stabilised within the binding pocket by means of π-interactions between the cyclohexene ring of bisabolol and Phe123, as well as π-interactions between aromatic ring of Trp114 and the unsaturated double bond of bisabolol. Further stabilisation is provided by numerous van der Waals contacts and hydrophobic interactions.

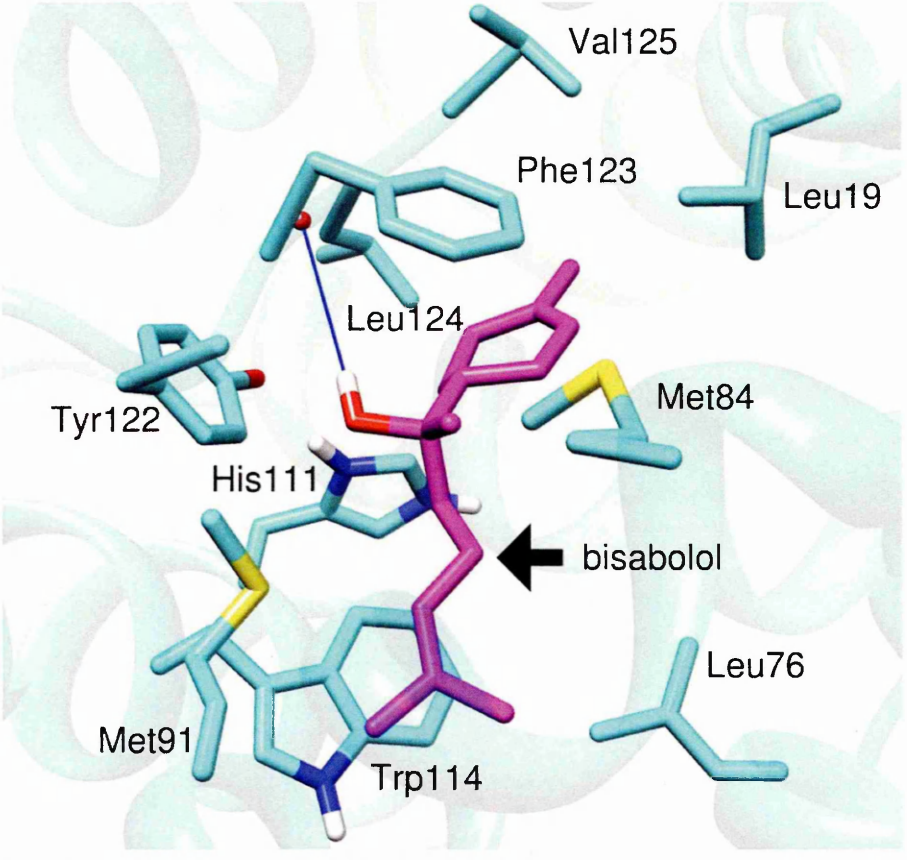



**Figure 4.7** ***Agam*OPB1: Binding site of bisabolol**
H-bond shown as a blue line

Docking of indole revealed the presence of a second binding pocket lying outside the main binding cavity of the OBP receptors. Indole binds in the second binding pocket with almost the same if not greater affinity (-5.4 kcal $mol^{-1}$ compared to -5.0 kcal $mol^{-1}$ in the main binding pocket). Figure 4.8a illustrates the binding of indole in the main and secondary binding pocket of *Agam*OPB1. The secondary binding pocket is surrounded by residues of the 3rd (Asn56, Phe59, His60) and 6th helices (Lys106, Pro108, Leu110), and loop residues Asp 70, Val71 lying between the 3rd

and 4th helices and Leu103 lying between the 5th and 6th helices. The bottom of this binding pocket is quite hydrophilic. To ascertain whether this pocket may function as a second binding site and, thus potentially play a biological role in the binding of indole, the Dowser program [198] was used to survey *Agam*OBP1 for internal cavities that are likely to be hydrated. The program identified two such sites one of which was coincident with the secondary binding site of the receptor as shown in Figure 4.8b thus putting into question the role of this cavity as a potential second site for the binding of indole.

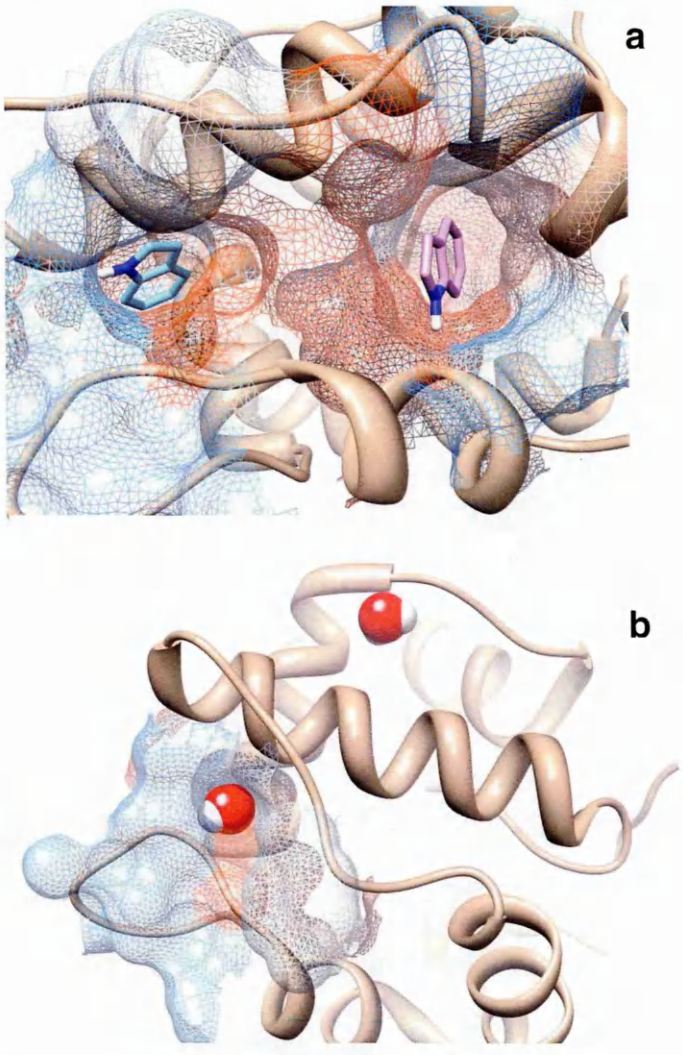



**Figure 4.8** ***Agam*OPB1: Binding site of indole**

**a.** The figure depicts indole in the main and secondary binding cavity; **b.** the figure shows that the secondary binding site is likely to be hydrophilic

The docked indole in the main pocket of *Agam*OBP1 is in the same relative position as DEET as shown in Figure 4.9. Overlaying the docked structure of the *Agam*OBP1-indole complex onto the X-ray model of *Agam*OBP1-DEET shows the benzene rings of indole and DEET to be almost coincident and be stabilised by means of π-π interactions with Trp114.

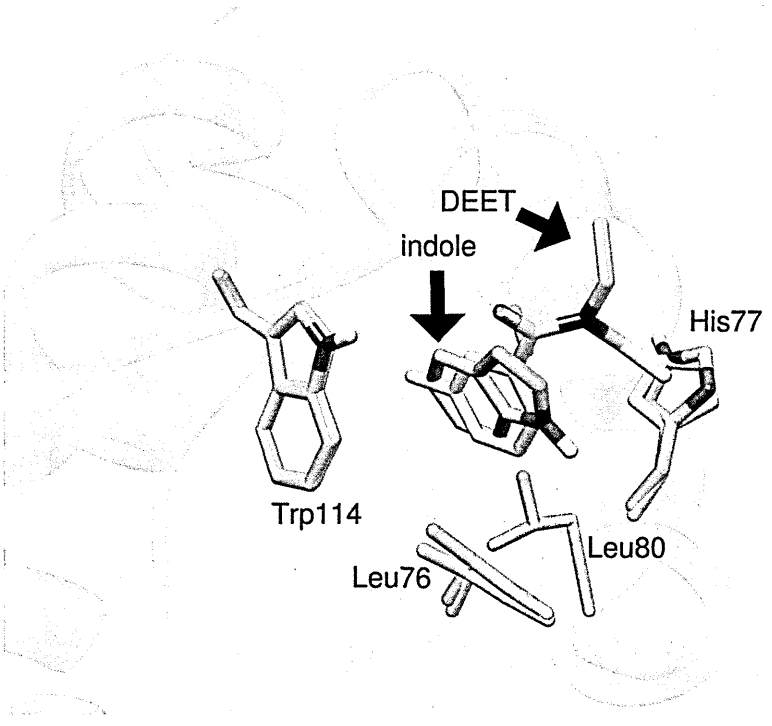



**Figure 4.9** **Relative positions of DEET and indole in the binding pocket of *Agam*OBP1**

The above docking simulations and the crystallographic models of *Agam*OBP1-DEET and *Cqui*OBP1-MOP present a rather intriguing picture. The X-ray models indicate that the these OBPs are able to bind ligands either at the interface between the two monomeric subunits (*Agam*OBP1-DEET) or, larger ligands, into a volume extending from the main binding cavity to the dimeric interface. Docking simulations, on the other hand, were able to reproduce the crystallographic models but, notwithstanding modelling artifacts, they also provided evidence that ligand binding can take place in the part of the binding cavity that is occluded by the C-terminus "lid" of the OBPs.

Hydrogen bond formation between the polar moieties of the ligands and the peptidic oxygen of Phe123 of *Agam*OBP1 (Phe122 of *Aaeg*OBP1 and *Cqui*OBP1) presents an interesting situation. In principle, given that the C-terminus is "fixed" onto helices 1 and 3 through interactions of the terminal carboxylate with the hydroxyl group of Tyr56 and the δ-N of His25 as described in Section 4.1.1, Figure 4.3, the substitution of Phe123(122) by any other amino acid residue should not affect ligand binding. Therefore, Phe123(122) cannot be considered to be a "selectivity filter" unless it functions as such in concert with other conserved residues such as Tyr122(121) and Trp114(113)[13]. As an example, in Figure 4.7, bisabolol is shown to be stabilised by π-interactions between Phe123 and the cyclohexene ring of the ligand, Trp114 and the unsaturated double bond of the aliphatic tail of the ligand, as well as numerous hydrophobic and van der Waals interactions involving Tyr122(121). This type of stabilisation is unlikely to occur with a ligand such as octanoic acid that is devoid of unsaturated double bonds and which was also shown to form H-bonds with Phe123 of *Agam*OBP1. However, it is possible that the -COOH group is involved in simultaneous H-bonding interactions, one with the peptidic oxygen of Phe123 in which it acts as a hydrogen donor and a second one with the amide hydrogen of Phe123 in which it acts as a H-bond acceptor. The simultaneous formation of H-bonds between ligands with appropriate functional groups, Tyr122 and Phe123 of the OBPs was observed on several cases involving the ligands shown in Figure 4.4 and, therefore, should be theoretically possible. Such H-bonding interactions are likely to keep the ligand "anchored" onto the C-terminus "lid" of the receptor and should this be proven to be the case experimentally, the C-terminus of the OBPs could be considered to constitute a selectivity filter for a specific type of odorant molecules.

---

[13] Residue numbers refer to *Agam*OBP1. Residue numbers in parentheses refer to *Aaeg*OBP1 and *Cqui*OBP1

The docking results presented above indicate that ligand binding in the main binding cavity or the tunnel leading to the interface between monomeric subunits depends on the fine interplay between hydrophobic, van der Waals and H-bonding interactions.

## 4.3.3 Docking on dimers

### 4.3.3.1 Background

Homo- and hetero-dimerisation of OBPs/PBPs under physiological conditions has been suggested on several occasions. The very high concentration (~ 10mM) of *A. polyphemus* PBP in the insect sensilla was taken as indication that, *in vivo,* it may be present as a homodimer [199]. The formation of PBP dimeric and higher-order multimeric forms has been demonstrated in electrophoretic experiments [21]. Furthermore, in ligand binding experiments, it was shown that a monomer/dimer equilibrium exists and that this duality may be relevant for the binding and release of pheromones [77].

Heteropolymerisation is also possible, as co-localisation of different PBPs has been demonstrated in some sensilla [84]. Evidence for heterodimer formation has arisen from co-immunoprecipitation and cross-linking studies involving different *Agam* OBPs [80]. Additional supporting evidence has been obtained from ligand-binding experiments. In recent fluorescent displacement assays, a number of odorant molecules were tested against six recombinant *A. gambiae* OBPs. It was found that for two of the OBPs, ligands bound with more than one binding constant suggesting the formation of heterodimers with different binding characteristics from those of the component proteins. The authors suggested that this phenomenon can be attributed to (1) dimer formation, (2) the degree of asymmetry of the dimer and (3) the dissociation constant for the equilibrium monomer/dimer [200]. These

findings are in line with earlier experiments involving the gypsy moth *Lymantria dispar* which demonstrated that low ligand:PBP ratios resulted in much weaker binding than assays performed at high ligand:PBP ratios. The authors suggested that multimers of PBP-ligand complexes may form at high pheromone concentration. This could have functional significance in that at low pheromone concentrations PBPs may act as pheromone transporters, whereas, at high concentrations, pheromones may induce PBP dimerisation and concomitant sequestration of excess pheromone thus preventing an overload of the receptors [16]. In the following section, the structural features of dimeric interfaces of mosquito OBP models are presented as well as the results of docking simulations performed in the space surrounding the binding interface. Most recently, a combination of crystallographic and NMR studies provided proof of the formation of a dimer between *Agam*OBP1 and *Agam*OBP4 upon binding of indole [194].

#### 4.3.3.2 Dimer intefaces

The data in Table 4.2 is derived from the Protein Interfaces, Surfaces and Assemblies (PISA) tool of the Protein Data Bank Europe [201]. It summarises some of the properties of the subunit interfaces of mosquito OBPs that have been crystallised in dimeric form.

| Dimer | Model | No. res | Int. Area (Å$^2$) | No.HB | No. SB | Score (CSS) |
|---|---|---|---|---|---|---|
| *Agam*OBP1 | 2erb | 16 | ~600 | 10 | 3 | 0.19 |
| | 3n7h | 15 | ~622 | 8 | 5 | 0.08 |
| *Aaeg*OBP1 | 3k1e | 15 | ~645 | 9 | 5 | 0.10 |
| *Cqui*OBP1 | 3ogn | 15 | ~640 | 8 | 6 | 0.01 |

**Table 4.2 Mosquito OBP protein-protein interfaces**

No. res. = number of interface residues per monomer; No. HB / SB = number of H-bonds and salt bridge; Int. Area = interface area per monomer

The largest interface area is observed in the case of *A. gambiae* OBP1. The solvation free energy gain upon formation of the interface ($\Delta^iG$) is negative indicating hydrophobic interfaces although of low absolute value. As a consequence the dimer interfaces have a low Complex Formation Significance Score (CSS <0.2) and therefore do not seem to play an essential role in complex formation. CSS ranges from 0 to 1 as the interface relevance to complex formation increases. The data implies that although there may not be a structurally conserved mechanism driving dimerisation, the latter would still possible if the concentration of the solute were high enough.

#### 4.3.3.3 Docking results

The docking simulation results, as well as the X-ray models of *Agam*OBP1-DEET and *Cqui*OBP1-MOP complexes, provide some evidence that the interface between the two monomeric subunits of the mosquito OBPs may have a role in the binding of odorants. Docking simulations were performed to ascertain the likelihood of ligand binding at the dimeric interface.

Affinity maps of 22x22x22 Å grid points and 0.375 Å spacing were centered at the interface between the two monomeric subunits as shown in Figure 4.10. All other parameters were set as described in Section 2.2.2.1.

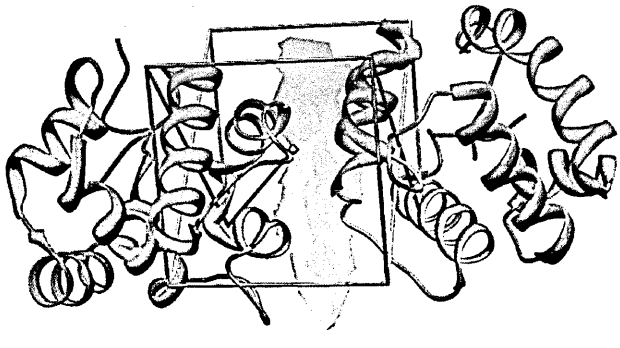


**Figure 4.10** **Location of the dimeric interface and grid box for docking simulations on OBP dimers**

The location of ligand binding at the dimeric interface of *Agam*OBP1 is exemplified in Figure 4.11.

**Figure 4.11** **Location of docked ligands at the dimer interface**

The free energies of binding of selected ligands in complex with *Agam*OBP1, *Aaeg*OBP1 and *Cqui*OBP1 are given in Table 4.3.

| | ***Agam*OBP1** | | ***Aaeg*OBP1** | | ***Cqui*OBP1** | |
|---|---|---|---|---|---|---|
| Ligand | $\Delta G_{dimer}$ | $\Delta G_{monomer}$ | $\Delta G_{dimer}$ | $\Delta G_{monomer}$ | $\Delta G_{dimer}$ | $\Delta G_{monomer}$ |
| indole | -5.1 | -5.5 | -5.2 | -5.1 | -5.2 | -5.5 |
| deet | -6.7 | -6.6 | -7.0 | -6.7 | -6.9 | -6.6 |
| geraniol | -5.5 | -5.4 | -5.7 | -5.7 | -5.6 | -5.4 |
| geranyl acetone | -6.6 | -6.7 | -6.9 | -7.0 | -6.8 | -6.5 |
| geranyl acetate | -6.1 | -6.2 | -6.3 | -6.4 | -6.4 | -6.1 |
| bisabolol | -8.2 | -7.7 | -8.3 | -8.4 | -8.2 | -7.6 |

**Table 4.3** **Free energies of binding of ligands in complex with mosquito OBP dimers and monomers**

A comparison of the binding affinities of the monomeric or dimeric states of the proteins towards the ligands show that they are almost identical despite the fact that the relative locations of ligand within the binding cavities of the proteins are different. Figure 4.12 provides an example of ligand docking locations in relation to the *Agam*OBP1 dimeric and monomeric forms. Much as in the case of ligand-protein monomer binding, hydrophobic interactions make the major contribution to ligand binding to the dimeric form of the proteins. As expected, the ligands make

extensive contacts with residues of α-helices 4 and 5 of both monomeric sub-units as well as Trp114 of the 6th α-helix. This residue, as well as His77 participate in non-polar, π-π interactions with the ligands. No H-bonding interactions were detected.

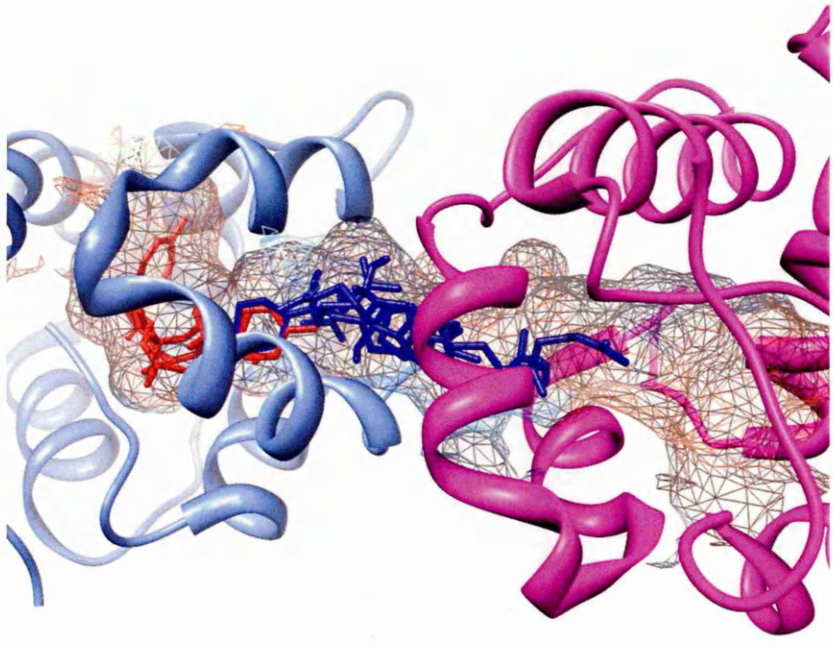


**Figure 4.12 Docked ligands relative to the monomeric and dimeric states of the protein**

Ligand docking poses in red colour show the results of docking in the main binding cavity of AgamOBP1. Ligands in blue colour show the results of docking at the interface between the two monomeric sub-units of the protein. Docking on the B-chain not shown

The figure above taken together with the results presented in Table 4.4 indicate that ligand binding may occur with the same affinity at the dimeric interface as well as within the main binding cavity of the proteins. However, the docking simulations performed in the course of this work did not take into account the presence of water molecules interacting with the binding partners. Such interactions have been shown in the crystallographic model of *Agam*OBP1-DEET [66]. There is another aspect that may influence binding at the dimeric interface, namely co-operativity of binding between two ligand molecules at each side of the dimeric interface. Some evidence in support of this has been suggested in the case of the *Cqui*OBP1-MOP complex [69].

## 4.4 MD simulations

For all MD simulations described henceforth the procedures followed for the MD simulations and for free energy calculations, per-residue free energy decomposition and trajectory analysis were as described in Sections 2.2.3.1, 2.2.5, 2.2.5.1 and 2.6, respectively. MD production simulations were carried out in explicit water with the PME method after an initial 2ns of unrestrained equilibration.

### 4.4.1 Binding of MOP

#### *4.4.1.1 Properties of the molecular systems used*

Table 4.4 provides a summary of the properties of the molecular systems used in MD simulations. For the *Cqui*OBP1-MOP simulation, the initial conformation of MOP was the same as that of the corresponding crystallographic model, whereas for the *Agam*OBP1- and *Aaeg*OBP1-MOP complexes the lowest energy conformation of MOP obtained from docking simulations were taken as starting conformations of the ligand. The α-chains of the dimeric proteins were taken as the starting conformation of the OBPs.

| | PDB id | No. aa | No. atoms | No. counterions | No. HOH | gyration radius (Å) | Box (Å) |
|---|---|---|---|---|---|---|---|
| *Agam*OBP1-MOP | 2erb | 123 | 1966 | 8 | 4081 | 13.61 | 69x69x69 |
| *Aaeg*OBP1-MOP | 3k1e | 124 | 1995 | 8 | 4193 | 13.65 | 69x69x69 |
| *Cqui*OBP1-MOP | 3ogn | 124 | 1979 | 7 | 3888 | 13.65 | 68x68x68 |

**Table 4.4** **MOP: Properties of molecular systems used in MD simulations**
No. aa & No. atoms are the number of residues and number of atoms of the protein; No. counterions is the number of ions used to neutralise the system; No. HOH is the number of water molecules contained in the solvation box

#### 4.4.1.2 Time-dependent properties

The crystallographic model (3ogn) was used as reference to validate the results of the MD simulation. The MD trajectory was subjected to B-factor analysis as this offers a convenient yardstick to compare the experimental and simulation results. It is noted that whereas the crystallographic B-factor includes not only thermal fluctuations but also contributions from the global translation and rotation of the protein, the B-factors calculated from the *Cqui*OBP1-MOP simulation were obtained by rigid body alignment and removal of the global translation and rotation. A comparison of the calculated B-factors against the X-ray model is shown in Figure 4.13.

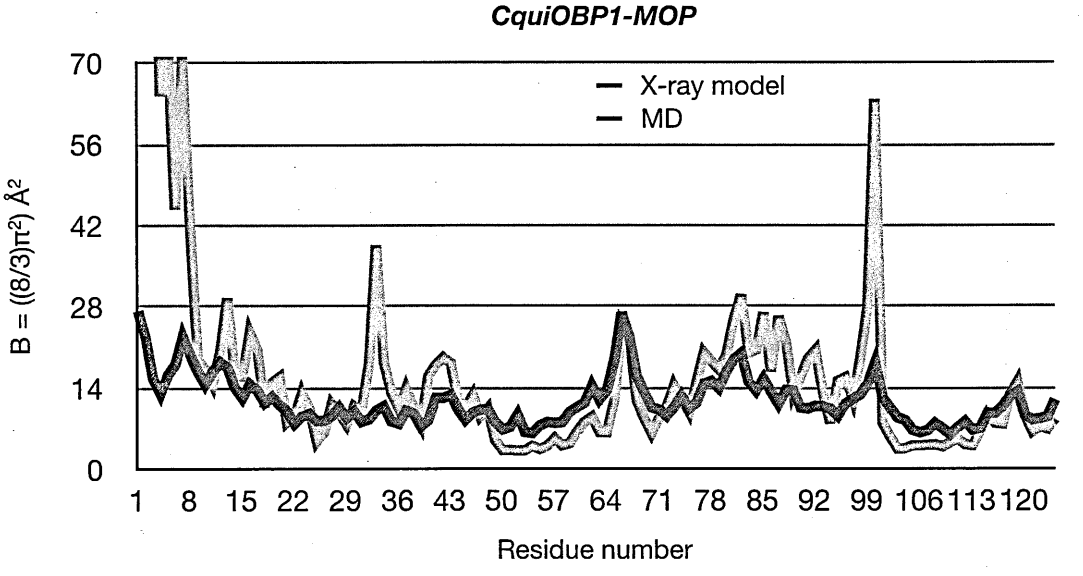



**Figure 4.13** **Comparison of calculated B-factors from the MD simulation and the X-ray model**

Horizontal axes are residue indices, and vertical axes are B-factors in Å²

In principle, the effect of rigid body fitting should give smaller B-factors than the experimental ones, other factors such as crystal packing may constrain the flexibility of residues involved in crystal contacts with a concomitant reduction in B-factors. Despite these differences, B-factor comparison provides a qualitative assessment of the simulation.

For each MD simulation, the range of RMSD of the OBP backbone atoms from the respective X-ray structures is given in Table 4.5.

| Ref. structure | complex | residues | avg[1] (Å) | sd[2] | min[3] (Å) | max[3] (Å) | Time (ns) |
|---|---|---|---|---|---|---|---|
| 2erb | *Agam*OPB1-MOP | 123 | 1.20 | 0.16 | 0.70 | 1.60 | 10 |
| 3k1e | *Aaeg*OBP1-MOP | 124 | 1.30 | 0.15 | 0.80 | 1.70 | 10 |
| 3ogn | *Cqui*OPB1-MOP | 124 | 1.30 | 0.22 | 0.80 | 1.90 | 10 |

**Table 4.5** **MOP: Backbone RMS deviation of MD simulation models**

1. average backbone RMSD during the trajectory with respect to the reference structure; 2. standard deviation; 3. minimum and maximum RMS deviation with respect to the reference structure

RMSD fluctuations of the protein backbones and MOP are shown in Figure 4.14. The RMSD of the protein Cα atoms is less that 2Å for the three OBPs. However that for *Cqui*OBP1 seems to be drifting indicating that the system is sampling different conformational spaces. The RMSD for MOP is higher in the *Agam*OBP1 and *Aaeg*OBP1 systems as compared to the *Cqui*OBP1-MOP complex. Figure 4.14 also shows H-bonds between MOP, the receptors and/or water molecules during the trajectories, and H-bond occupancies are given in Table 4.6.

| | **VMD[1]** | | | **cpptraj[1]** | | |
|---|---|---|---|---|---|---|
| | Donor | Acceptor | Occupancy (%) | Donor | Acceptor | Occupancy (%) |
| *Agam*OBP1 | WAT@H | MOP@O4 | 0.1 | WAT@H | MOP@O4 | < 0.1 |
| | Ser77@HG | MOP@O4 | 0.2 | His109@HE2 | MOP@O4 | < 0.1 |
| | | | | Ser77@HG | MOP@O4 | < 0.1 |
| *Aaeg*OBP1 | His110@HE2 | MOP@O4 | 4.4 | His110@HE2 | MOP@O4 | 30 |
| *Cqui*OBP1 | WAT@H | MOP@OAV | 5.5 | WAT@H | MOP@OAV | 13 |

**Table 4.6** **H-bond formation between the *Agam*OBP1, *Aaeg*OBP1, *Cqui*OBP1 and MOP**

[1] See Section 2.2.6.2 H-bond formation

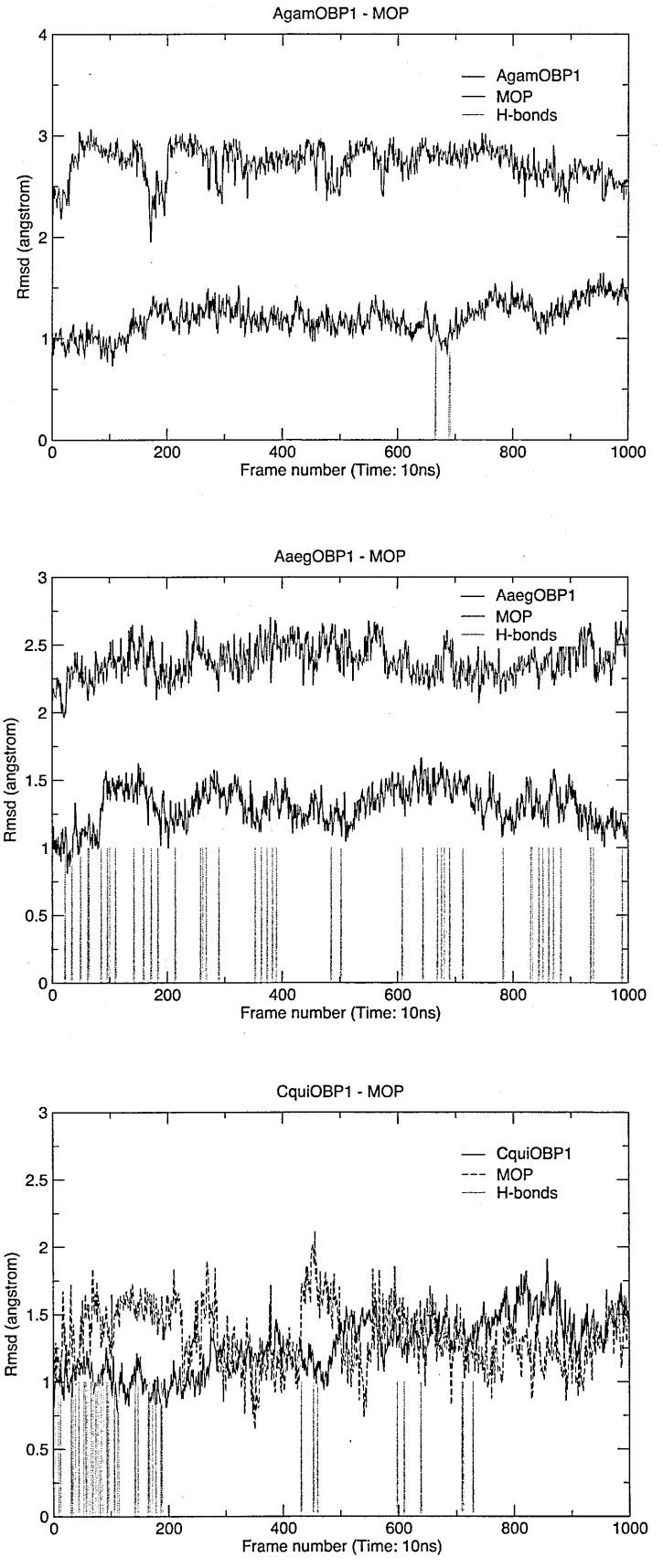



**Figure 4.14** **RMSD time series of mosquito OBP Cα atoms and MOP from the starting structures**

Although the oxygen atoms of the lactone ring and the acetyl ester group can function as proton acceptors, no H-bonds were detected in the crystallographic study, in which the authors also found that despite the dissociation of the dimer in solution, binding was retained [69]. In the MD simulations, H-bonds formation was transient for the *Agam*OBP1- and *Cqui*OBP1-MOP complexes but not so for the *Aaeg*OBP1-MOP complex. The occupancy of the H-bond between His110-NE2 of *Aaeg*OBP1 acting as a hydrogen donor and the carbonyl oxygen of the MOP lactone ring acting as acceptor was in the range of 4.4 - 30% depending on the angle cutoff values of the detection method used.

During the trajectory, MOP was in contact with numerous residues within the binding cavities of the receptors, mainly through hydrophobic and van der Waals interactions with the exception of His110 of *Agam*OBP1 which interacted with MOP electrostatically. Residues that made the strongest contribution to the free energies of binding (< -7.0 kcal $mol^{-1}$) were identical for A*aeg*OBP1 and *Cqui*OBP1 and differed to some extent in the case of *Agam*OBP1. Most of these residues belong to the 4th and 5th helices which together with Trp116 of the 6th helix form the tunnel lying at the interface between the two monomeric subunits.

The trajectories were subjected to clustering analysis using the average linkage algorithm (Section 2.2.6.1) in order to investigate whether there were any major conformational transitions during the course of the MD simulations. Conformations in each trajectory were grouped into five different clusters which were then analysed in terms of structural features and free energies of binding. The crystallographic structure (3ogn) was used as reference to ascertain whether the conformational space sampled in the course of the trajectories included similar protein-ligand conformations.

Superposition of representative structures from each of the five clusters of the *Cqui*OBP1- MOP trajectory onto the X-ray structure shows that ligands are found in the same relative position within the binding pocket and adopt very similar poses when compared to the ligand in the crystallographic structure. The lactone head of the ligand lies well within the binding cavity and, in the given MD simulation, it appears able to rotate by approximately 45° with respect to the plane of the lactone ring of the crystallographic model as shown in Figure 4.15. In four out of the five clusters, the lactone rings of the representative structures were almost coincident with that of the X-ray model.

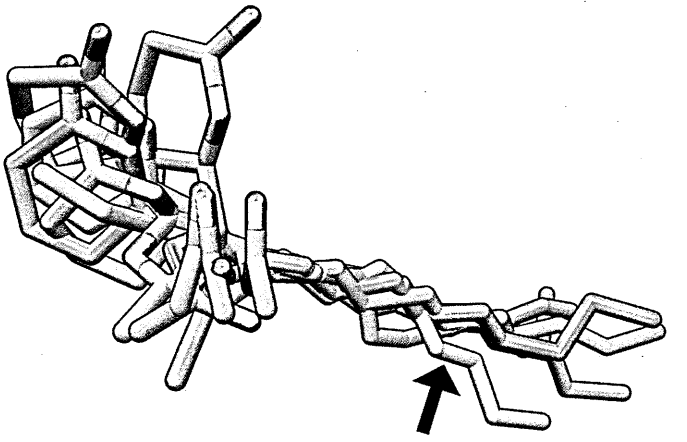



**Figure 4.15** **Representative structures of MOP obtained from MD simulation juxtaposed on the X-ray model**

As mentioned in Section 4.4.1.1, the initial conformation of MOP at the start of the simulation time was the same as that of the corresponding crystallographic model and the molecule did not show any major translational movement during the trajectory. The rotation of the lactone ring within the binding pocket is in agreement with the X-ray model in which the B-factor of the lactone head of MOP was found to be considerably higher than that of the protein backbone atoms. Figure 4.16 shows that relative orientation of MOP in the *Agam*OBP1- and *Aaeg*OBP1-MOP complexes relative to that in crystallographic *Cqui*OBP1-MOP complex. It can be

seen that MOP in the *Agam*OPB1 complex is almost coincident to MOP in the *Cqui*OBP1-MOP complex contrary to MOP in the *Aaeg*OBP1 complex the lactone head of which is at 90° to that of MOP in the *Cqui*OBP1-MOP complex.

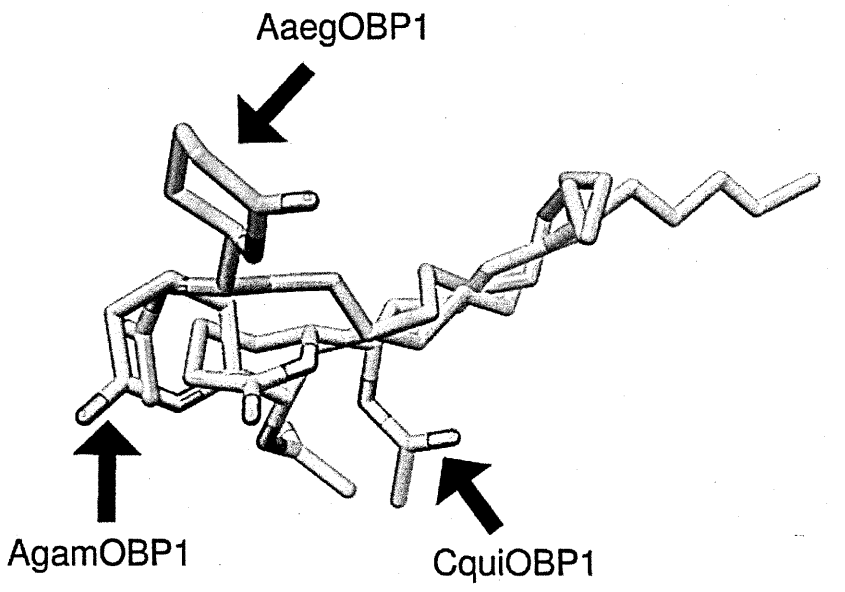



**Figure 4.16** **Relative orientations of MOP at the start of the MD simulations**
The arrows show MOP in the corresponding mosquito OBP

Superposition of the representative structures from the minimum energy clusters of the *Agam*OPB1- and *Aaeg*OBP1-MOP complexes onto the crystallographic model of *Cqui*OBP1 (3ogn) shows that the hydrophobic tail of MOP of the *Agam*OBP1-MOP complex lies well within the binding pocket of the receptor unlike the *Aaeg*OBP1-MOP and *Cqui*OBP1-MOP models in which the ligand tail protrudes out of the binding pocket into the dimer interface (Figure 4.17). A comparison of Figures 4.16 and 4.17 shows that MOP from a "compact" initial position in *Aaeg*OBP1 adopts, at the end of the simulation, a stretched conformation with the lactone head of the molecule lying on the same plane with its tail end, similar to the crystallographic model of *Cqui*OBP1. The opposite change in conformation and relative orientation of MOP is observed in the *Agam*OBP1 simulation.

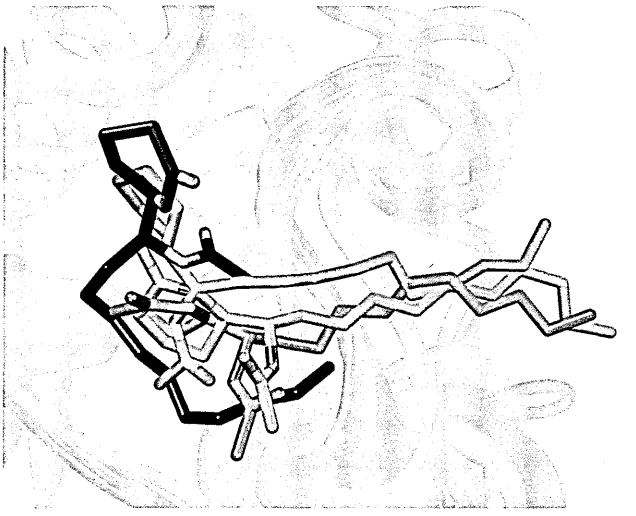


**Figure 4.17** **Superposition of the minimum energy structures obtained from the mosquito OBP1-MOP 10ns MD trajectories onto the X-ray model of *Cqui*OBP1-MOP**

magenta = *Cqui*OBP1-3OG (X-ray model); blue = *Agam*OBP1-3OG; green = *Aaeg*OBP1-3OG; cyan = *Cqui*OBP1-3OG

Although much longer sampling of the conformational space would be needed before any definitive conclusions are drawn regarding optimal binding, the observed changes of orientation and relative location of MOP in the three binding sites may suggest some fine differences in binding specificity of the OBPs.

### 4.4.1.3 Time-independent properties: binding free energy components

The free energies of binding were subjected to free energy decomposition. Free energy decomposition was also performed on the lowest energy clusters (see Table 4.7) resulting from clustering analysis of the MD trajectories.

| ***Agam*OBP1-MOP** | | ***Aaeg*OBP1-MOP** | | ***Cqui*OBP1-MOP** | |
|---|---|---|---|---|---|
| Occupancy (%) | $\Delta G$ (kcalmol$^{-1}$) | Occupancy (%) | $\Delta G$ (kcalmol$^{-1}$) | Occupancy (%) | $\Delta G$ (kcalmol$^{-1}$) |
| 13.8 | -37.6 | 9.0 | -40.9 | 48.1 | -41.1 |
| 41.2 | -37.1 | 10.3 | -41.2 | 0.02 | -38.7 |
| 14.9 | -38.2 | 50.3 | -41.6 | 36.3 | -40.4 |
| 20.5 | -36.8 | 3.1 | -40.2 | 4.6 | -41.2 |
| 9.6 | -38.2 | 27.3 | -40.6 | 10.8 | -41.8 |

**Table 4.7** **Clustering results of *Agam*OBP1-, *Aaeg*OBP1- and *Cqui*OBP1-MOP MD 10ns trajectories**

The results presented in Table 4.8 show that the free energies of binding for the *Aaeg*OBP1- and *Cqui*OBP1-MOP complexes are very similar and lower by over 2 kcalmol$^{-1}$ compared to that of the *Agam*OBP1-MOP complex. This is indicative of much higher binding affinities for *Aaeg*OBP1 and *Cqui*OBP1 towards MOP than for *Agam*OBP1. The total free energies of binding of the *Aaeg*OBP1 and *Cqui*OBP1 complexes are very similar but due to different contributing components. *Aaeg*OBP1 shows stronger electrostatic interactions with MOP in the range of 12 kcal mol$^{-1}$, which are counterbalanced by a solvation free energy penalty of almost equal magnitude to the total free energy of binding of the *Cqui*OBP1-MOP complex.

| | ***Agam*OBP1** | | ***Aaeg*OBP1** | | ***Cqui*OBP1** | |
|---|---|---|---|---|---|---|
| | 10 ns MD | Cluster 5 | 10 ns MD | Cluster 3 | 10 ns MD | Cluster 5 |
| Contrib[a] | Δ value[b] | Δ value[b] | Δ value[b] | Δ value[b] | Δ value[b] | Δ value[b] |
| $\Delta H_{vdW}$ | -48.2 | -49.2 | -50.2 | -50.7 | -50.3 | -50.3 |
| $\Delta H_{elec}$ | -2.0 | -1.9 | -12.7 | -12.0 | -0.1 | -0.6 |
| $\Delta H_{gas}$ | -50.2 | -51.1 | -62.9 | -62.7 | -50.4 | -50.9 |
| $\Delta G_{GB}$ | 18.3 | 19.0 | 28.0 | 27.5 | 15.7 | 15.0 |
| $\Delta G_{np}$ | -5.9 | -6.1 | -6.4 | -6.5 | -6.1 | -6.0 |
| $\Delta G_{solv}$ | 12.5 | 12.9 | 21.6 | 21.0 | 9.5 | 9.1 |
| $\Delta G_{gas+solv}$ | -37.7 | -38.2 | -41.3 | -41.6 | -40.9 | -41.8 |
| $\Delta TS_{total}$ | -23.5 | | -24.6 | | -24.6 | |
| $\Delta G_{total}$ | -14.2 | | -16.7 | | -16.3 | |

**Table 4.8** **Binding free energies and free energy components of mosquito OBP-MOP complexes**

Notation with regard to superscripts [a] and [b] is given in Table 3.15, p. 138

In view of the high sequence identity which also pertains to binding site residues this difference in electrostatic interactions between the *Aaeg*OBP1-MOP and *Cqui*OBP1-MOP complexes is interesting. Per-residue energy decomposition reveals differences in binding site interactions in some detail.

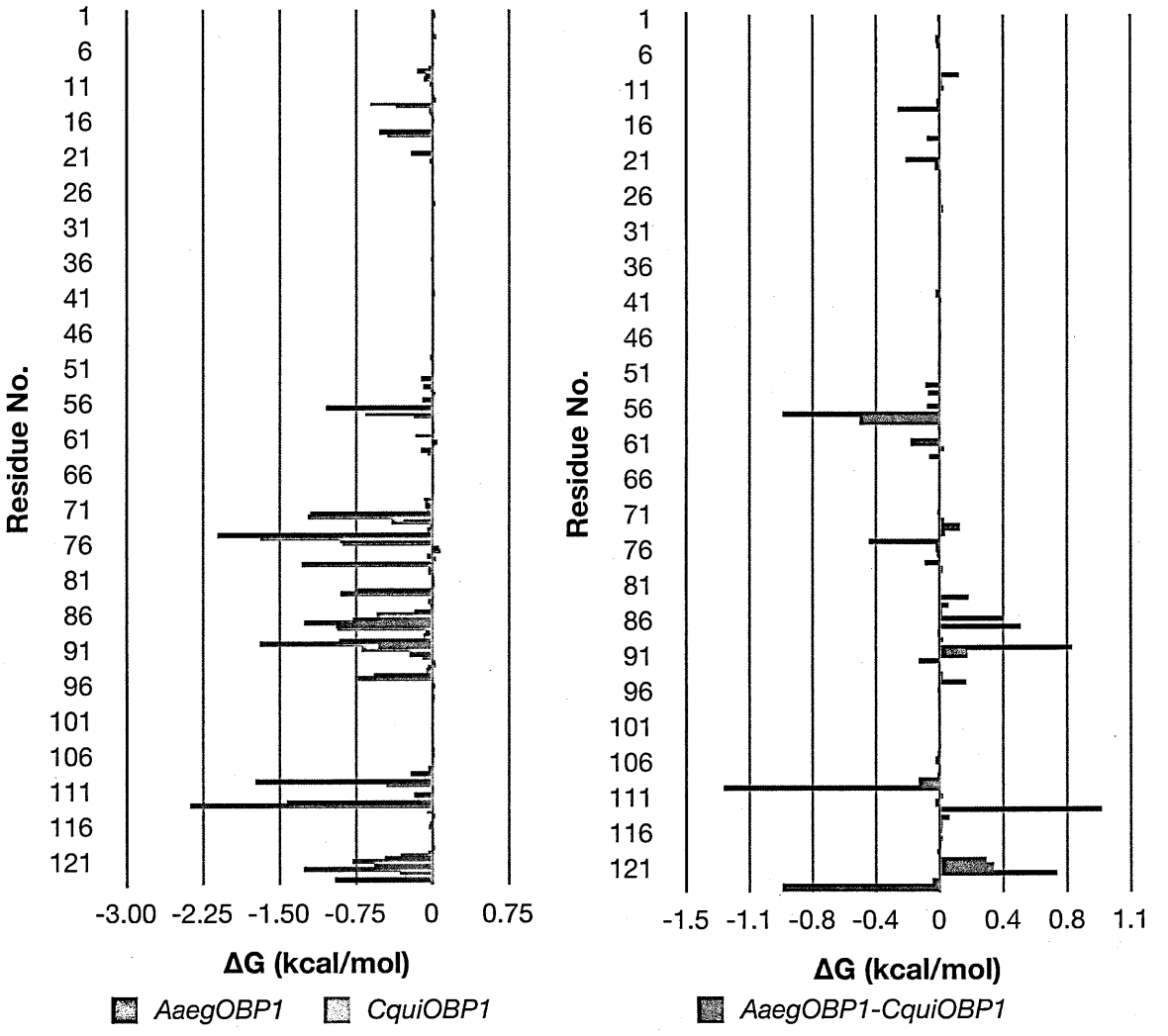



**Figure 4.18** **Per-residue decomposition of "effective energies" of binding[14] of the *Aaeg*OBP1- and *Cqui*OBP1-MOP complexes**

## 4.4.2 Binding of *N,N*-diethyl-*m*-toluamide (DEET)

### 4.4.2.1 Properties of the molecular systems used

Table 4.9 provides a summary of the properties of the molecular systems used in MD simulations. Two independent MD simulations were performed for each OBP-DEET complex. For one of these (*Agam*OBP1$_{[b]}$) the starting position of DEET was the same as that of the crystallographic model (3n7h). In all the other simulations, the starting conformations of DEET relative to the receptors were taken from different poses of the lowest energy docking clusters. The α-chains of the dimeric

[14] "Effective energy" refers to $\Delta G_{gas+solv}$

proteins were taken as the starting conformation of the OBPs. The boxed subscripts [a] and [b] after the name of the OBPs indicate one of the two independent MD simulations.

| | PDB id | No. aa | No. atoms | No. counterions | No. HOH | Box (Å) |
|---|---|---|---|---|---|---|
| *Agam*OBP1[a]-DEET | 3n7h | 125 | 1992 | 9 | 3782 | 68x68x68 |
| *Agam*OBP1[b]-DEET | 3n7h | 125 | 1992 | 9 | 3747 | 68x68x68 |
| *Aaeg*OBP1[a]-DEET | 3k1e | 124 | 1995 | 8 | 4281 | 70x70x70 |
| *Aaeg*OBP1[b]-DEET | 3k1e | 124 | 1995 | 8 | 4311 | 70x70x70 |
| *Cqui*OBP1 [a]-DEET | 3ogn | 124 | 1979 | 7 | 3936 | 68x68x68 |
| *Cqui*OBP1[b]-DEET | 3ogn | 124 | 1979 | 7 | 3929 | 68x68x68 |

**Table 4.9** **DEET: Properties of molecular systems used in MD simulations**

No. aa & No. atoms are the number of residues and number of atoms of the protein; No. counterions is the number of ions used to neutralise the system; No. HOH is the number of water molecules contained in the solvation box

### 4.4.2.2 Time-dependent properties

For each trajectory, the range of RMSD of the OBP backbone Cα atoms from the respective X-ray structures is given in Table 4.10 and the time evolution of the RMSD in Figure 4.19. In all cases the difference between the minima and the maxima of RMSDs with respect to the starting structures were less than 1.1 Å.

| Ref. structure[1] | complex | residues | avg[2] (Å) | sd[3] | min[4] (Å) | max[4] (Å) | Time (ns) |
|---|---|---|---|---|---|---|---|
| 3n7h | *Agam*OBP1[a]-DEET | 125 | 1.50 | 0.17 | 0.90 | 2.00 | 20 |
| | *Agam*OBP1[b]-DEET | 125 | 1.30 | 0.16 | 0.80 | 1.80 | 10 |
| 3k1e | *Aaeg*OBP1[a]-DEET | 124 | 1.10 | 0.13 | 0.70 | 1.50 | 10 |
| | *Aaeg*OBP1[b]-DEET | 124 | 1.30 | 0.13 | 0.90 | 1.70 | 10 |
| 3ogn | *Cqui*OBP1 [a]-DEET | 124 | 1.00 | 0.11 | 0.70 | 1.50 | 10 |
| | *Cqui*OBP1[b]-DEET | 124 | 1.20 | 0.14 | 0.80 | 1.60 | 10 |

**Table 4.10** **DEET: Backbone RMS deviation of MD simulation models**

1. average backbone RMSD during the trajectory with respect to the reference structure; 2. standard deviation; 3. minimum and maximum RMS deviation with respect to the reference structure

### *Agam*OBP1-DEET

The time evolution of the RMSD of the Cα atoms of the protein backbone is shown in Figure 4.19 on page 180 together with the interatomic distances between the amide oxygen of DEET (depicted as DEET@O), the backbone oxygen atom of Gly92 (depicted as Gly92@O) and the backbone oxygen of Phe123 (depicted as Phe123@O). The particular atoms were chosen as they have been shown to be involved in H-bonding and/or polar interactions in docking simulations. A similar notation (@) applies for other atomic designations[15]. Table 4.11 gives the occupancies of H-bonds over the MD simulation time.

| | VMD[1] | | | cpptraj[1] | | |
|---|---|---|---|---|---|---|
| | **Donor** | **Acceptor** | **Occupancy (%)** | **Donor** | **Acceptor** | **Occupancy (%)** |
| *Agam*OBP1[a] | Phe123@NH | DEET@O | 0.05 | Gly92@NH | DEET@O | 0.04 |
| | Gly92@NH | DEET@O | 0.1 | Phe123@NH | | < 1.0 |
| | WAT@H | DEET@O | 0.05 | WAT@H | | < 1.0 |
| *Agam*OBP1[b] | WAT@H | DEET@O | 56.0 | WAT@H1 | DEET@O | 50.0 |

**Table 4.11** **H-bond formation between the *Agam*OBP1 and DEET**

[1] See Section 2.2.6.1 H-bond formation

Although the amide oxygen of DEET is a proton acceptor, H-bonds were observed only transiently. Figure 4.18 and Table 4.11 show that effectively, no H-bonds were formed between DEET and the receptor during the *Agam*OBP1[a]-DEET trajectory. The figure also shows an upward drift of the RMSD of the protein backbone Cα atoms during the first 5ns from the start of the MD simulation after which it stabilised. At ~5ns from the start of the simulation, DEET underwent a sharp conformational change and an abrupt transition in the interatomic distances between DEET@O, Phe123@N and Gly92@N bringing the ligand in close

[15] This notation applies throughout this Chapter

proximity to Gly92 while moving it away from Phe123. Figure 4.19 shows that the initial position of DEET@O was ~4 Å from Phe123@N and ~8.5 Å from Gly92@N. DEET remained in the vicinity of Trp112 and Gly92 throughout the rest of the simulation time. It is recalled from Section 4.1.1.1 that the X-ray model of *Agam*OPB1-DEET shows H-bonding between the amide oxygen of DEET and a water molecule, which in turn interacts with the peptidic oxygen of Gly92 and with the NE1 nitrogen atom of Trp114. It is recalled that Phe123 is located in the so-called "lid" formed by the C-terminus of the proteins, whereas Gly92 is located at the interface between the monomeric subunits.

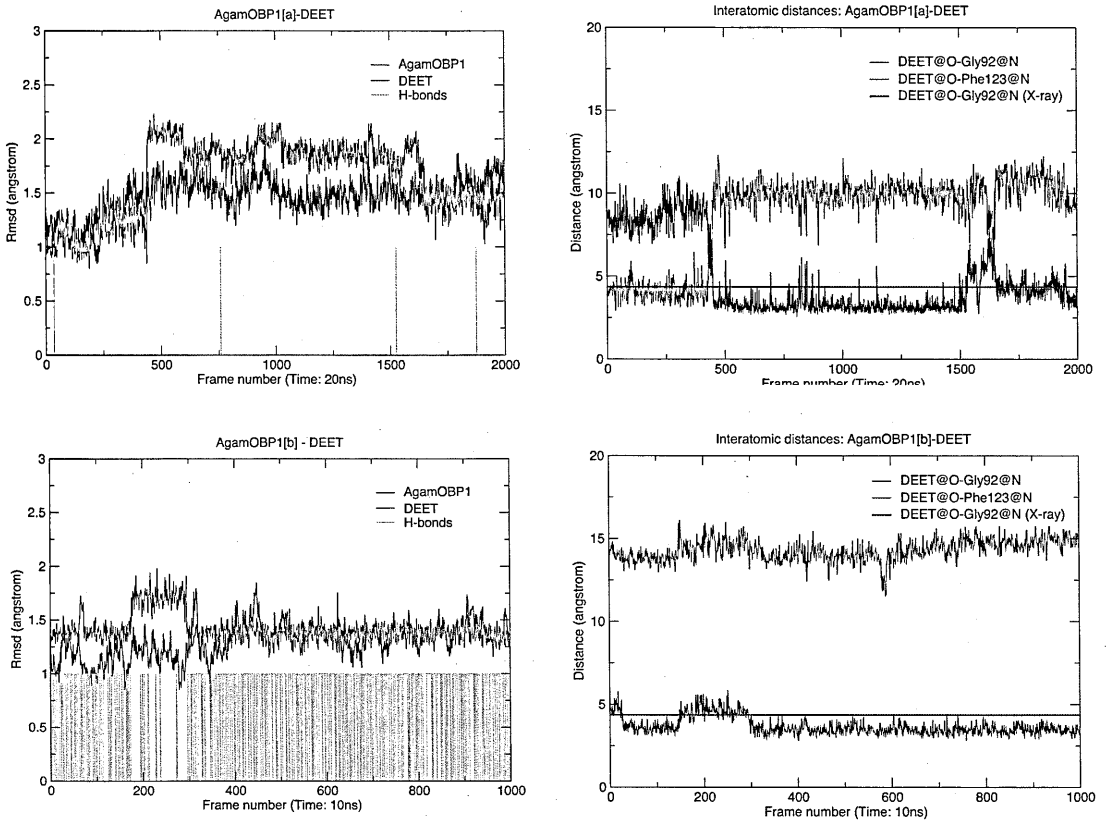



**Figure 4.19** **RMSD time series of *Agam*OBP1 Cα atoms and DEET from the starting structures**

In the *Agam*OBP1$_{[b]}$-DEET MD simulation, the starting position of DEET with respect to Gly92@N and Phe123@N was approximately that of the X-ray model and it remained so throughout the trajectory. Extensive H-bond formation between DEET and water was observed throughout the MD simulation.

The trajectories were subjected to clustering analysis using the average linkage algorithm (Section 2.2.6.1). Representative structures from the lowest energy clusters of *Agam*OBP1$_{a]}$-DEET and *Agam*OBP1$_{[b]}$-DEET superimposed over the crystallographic model of *Agam*OBP1-DEET are shown in Figure 4.20.

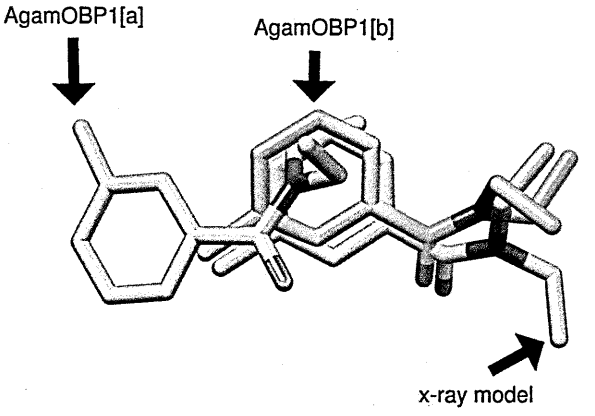



**Figure 4.20** **Superposition of representative low energy MD structures onto the X-ray model of *Agam*OBP1-DEET complex, showing only the ligand**

The figure shows the amide oxygen of DEET of the low energy representative structures from the two MD simulations to be equidistant to the peptidic oxygen of Gly92. Furthermore, whereas the benzene ring of DEET in the *Agam*OBP1$_{[b]}$-DEET simulation is almost at the same position as that of the X-ray model, it is not so in the *Agam*OBP1$_{[a]}$-DEET simulation. The possibility that this may be due to the two MD simulations sampling different conformational spaces should not be excluded. However, it is also likely that this is due to H-bonding interactions with water molecules. As mentioned earlier (Section 4.1.1.1), an H-bonding interaction network between the amide oxygen of DEET, a water molecule, the peptidic oxygen of Gly92 and the NE1 atom of Trp114 has been observed in the X-ray *Agam*OBP1-DEET model. In one of the two MD simulations (*Aaeg*OBP1$_{[a]}$-DEET), H-bonding interactions between water molecules Gly92 and Trp114 were shown for ~2% and ~4% of the simulation time, respectively. The corresponding H-bond occupancies in the *Agam*OBP1$_{[b]}$-DEET simulation were ~2% and ~78%, respectively.

### *Aaeg*OBP1-DEET

The time evolution of the RMSD of the Cα atoms of the protein backbone is shown in Figure 4.21 together with the interatomic distances between DEET@O, Gly91@O and Phe121@O and the observed H-bonds between DEET, protein residues and water molecules, details of which are given in Table 4.12.

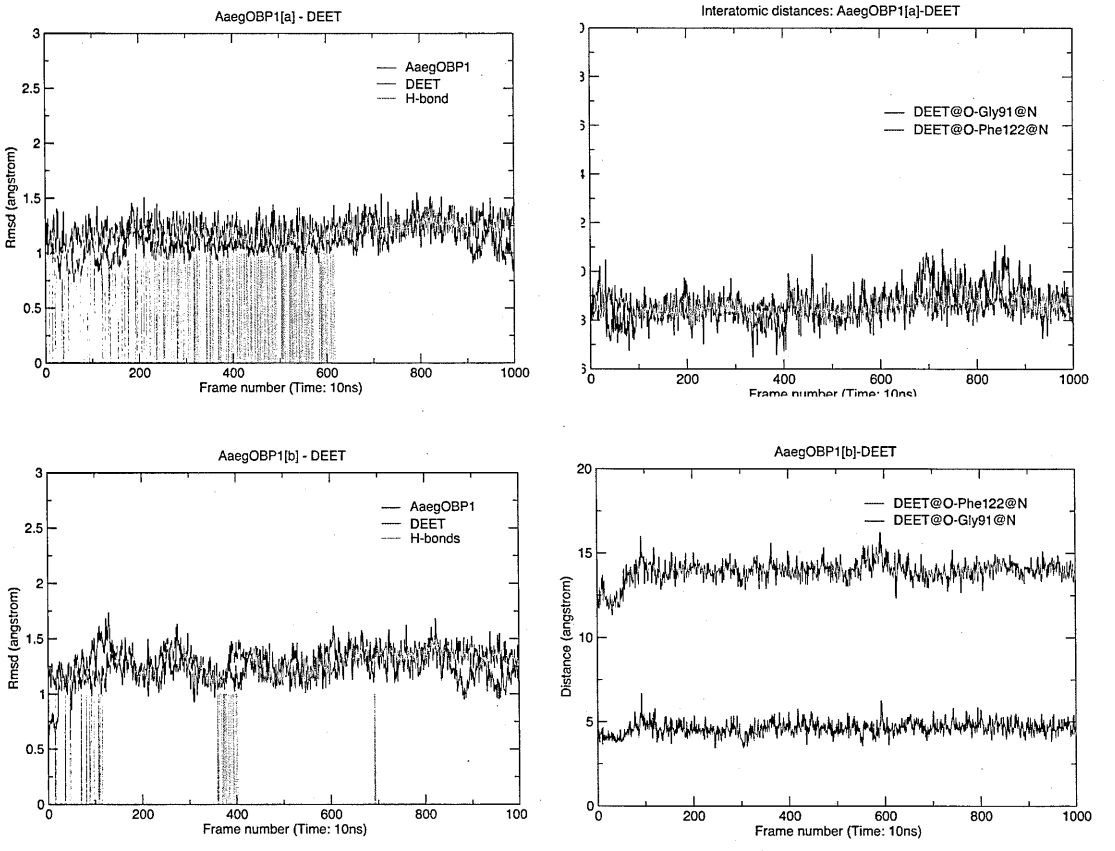



**Figure 4.21** **RMSD time series of A*aeg*OBP1 Cα atoms and DEET from the starting structures**

The figure shows the starting position of the amide oxygen of DEET which, for *Aaeg*OBP1$_{[a]}$-DEET, is almost equidistant to the amide nitrogens of Gly91 and Phe122 at the start of the MD simulation. There were no major fluctuations in the RMSD of the protein backbone atoms or the ligand and very little change in the relative interatomic distances during the trajectory. H-bond formation between DEET and water molecules exclusively was observed during the first 6ns of the MD simulation.

| | VMD[1] | | | cpptraj[1] | | |
|---|---|---|---|---|---|---|
| | **Donor** | **Acceptor** | **Occupancy (%)** | **Donor** | **Acceptor** | **Occupancy (%)** |
| *Aaeg*OBP1$_{[a]}$ | WAT@H | DEET@O | 19.7 | WAT@H1 | DEET@O | 22.7 |
| *Aaeg*OBP1$_{[b]}$ | TRP@NE1 | DEET@O1 | 0.4 | TRP@HE1 | DEET@O | 2.0 |
| | WAT@O | DEET@O1 | 3.1 | WAT@H1 | DEET@O | 3.0 |

**Table 4.12** **H-bond formation between the *Aaeg*OBP1 and DEET**

[1] See Section 2.2.6.2 H-bond formation

In the second MD simulation (*Aaeg*OBP1$_{[b]}$-DEET) the starting interatomic distances were ~4.5 Å and ~12.5 Å for DEET@O-Gly91@N and DEET@O-Phe122@N, respectively. The RMSD of the protein Cα atoms showed fluctuations within 1 Å and there was no significant change from the starting interatomic distances during the MD simulation time. No H-bond formation involving DEET was observed.

Following clustering analysis the results of which are presented in (Supplementary Material: *Agam*OBP1, *Aaeg*OBP1, *Cqui*OBP1 simulations: Clustering Results), representative structures from the lowest energy clusters of the two trajectories were juxtaposed against the X-ray model of *Agam*OBP1 for comparison with regard to the relative location of binding. Figure 4.22 shows the representative structure of DEET from cluster 1 of the *Aaeg*OBP1$_{[b]}$-DEET being in an almost identical location as in the *Agam*OBP1-DEET crystallographic model. It also shows the relative position of DEET from cluster 5 of *Aaeg*OBP1$_{[a]}$-DEET trajectory.

Figures 4.21 and 4.22 show the amide oxygen of DEET at the start of the simulation time being in close proximity to Trp113 with which it formed H-bonds transiently. During this trajectory the rotational and translational movement of DEET in the binding cavity was insignificant due to possible stabilisation through

interactions with Trp113 and "entrapment" of DEET in a low energy conformational landscape.

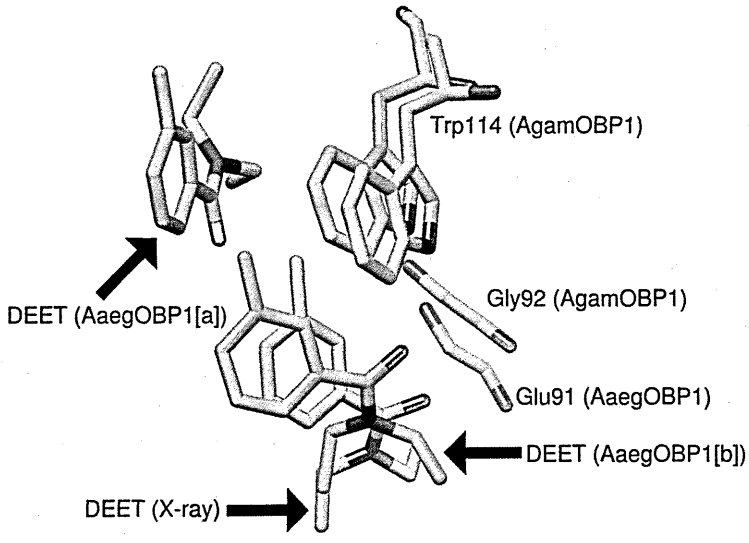



**Figure 4.22** **Superposition of representative low energy MD *Aaeg*OBP1-DEET structures onto the X-ray model of *Agam*OBP1-DEET complex, showing only the ligand**
Trp114 of *Agam*OBP1 corresponds to Trp13 of *Aaeg*OBP1

### *Cqui*OBP1-DEET

The time evolution of the RMSD of the Cα atoms of the protein backbone is shown in Figure 4.23 together with the interatomic distances between DEET@O, Gly91@O and Phe121@O. No H-bonds were observed between DEET, protein residues and water molecules. The figure shows that in the first MD simulation (*Cqui*OBP1$_{[a]}$-DEET), neither the RMSD of the protein Cα atoms or DEET fluctuated by more than 1 Å with respect to the initial structure. The interatomic distances between DEET@O, Gly91@N and Phe122@N remained relatively unchanged during the simulation time. In the second MD simulation (*Cqui*OBP1$_{[b]}$-DEET), the RMSD of the Cα atoms of *Cqui*OBP1 shows an upward drift during the first 3ns from the start of the MD simulation and a significant change of the RMSD of DEET. For the remainder of the simulation time the RMSD of the protein backbone atoms and DEET remained stable. The interatomic distances DEET@O-Gly91@N and DEET-Phe122@N showed little change during the trajectory.

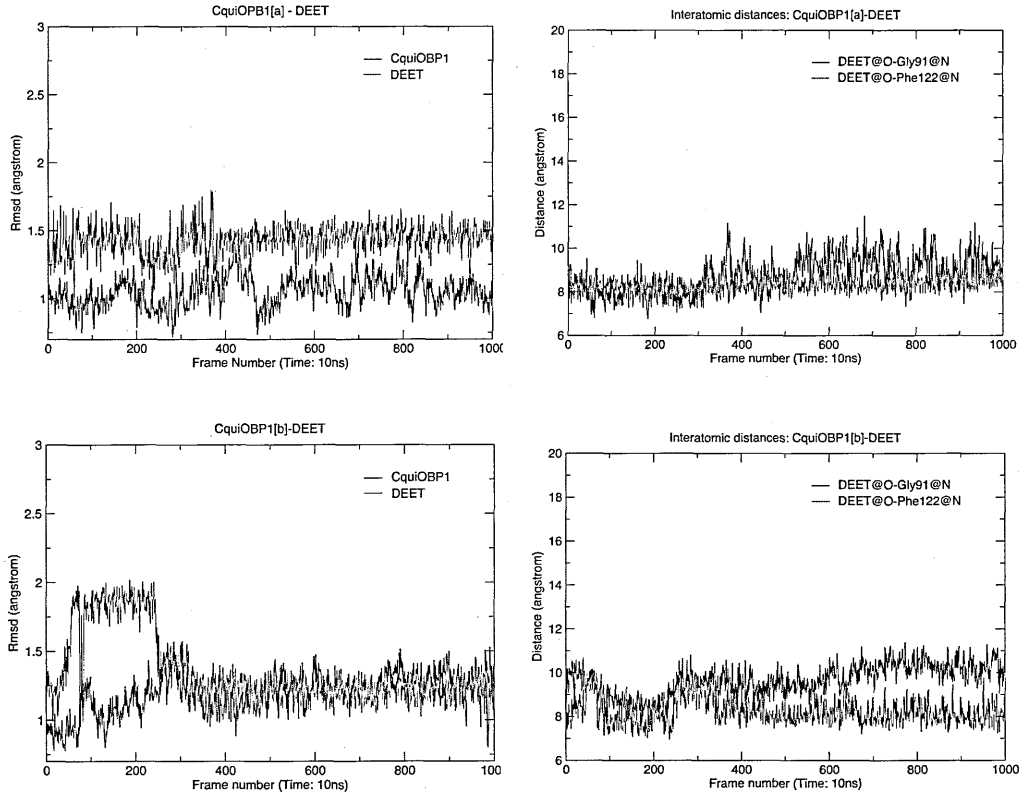



**Figure 4.23** **RMSD time series of *Cqui*OBP1 Cα atoms and DEET from the starting structures**

The positions of DEET from representative structures from the clusters with the lowest free energy of binding of each trajectory relative to DEET in the *Agam*OBP1-DEET complex are shown in Figure 4.24.

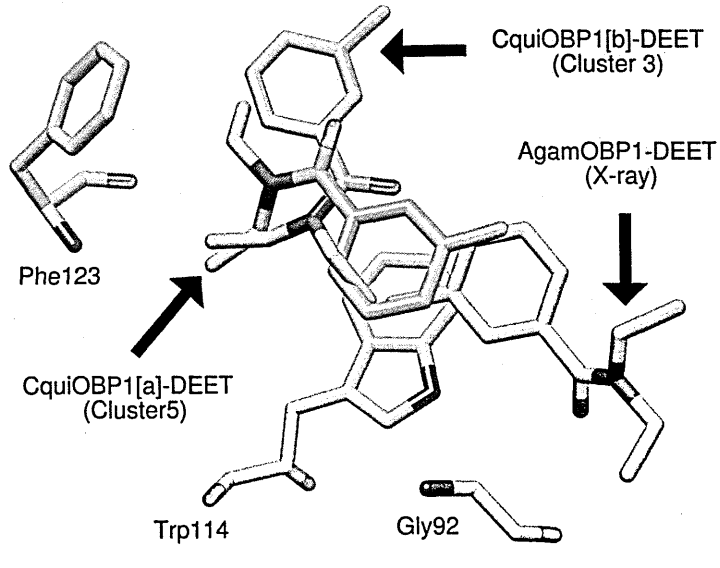



**Figure 4.24** **Representative low energy *Cqui*OBP-DEET structures juxtaposed against the X-ray model of *Agam*OBP1-DEET complex**

The figure above shows that although the benzene ring of DEET in *Cqui*OBP1$_{[b]}$-DEET simulations tends to the position of the X-ray model of *Agam*OBP1-DEET, the amide groups of the two structures are in anti-diametric positions. A longer trajectory time and/or repeat trajectories from different starting conformations would be required to generate a better picture of optimal binding.

#### *4.4.2.3 Time-independent properties*

**Binding free energy components**

The free energies of binding obtained from the individual MD simulations are summarised in Table 4.13.

| | *Agam*OBP1 [a] | *Agam*OBP1 [b] | *Aaeg*OBP1 [a] | *Aaeg*OBP1 [b] | *Cqui*OBP1 [a] | *Cqui*OBP1 [b] |
|---|---|---|---|---|---|---|
| ΔG (kcalmol$^{-1}$) | -25.3 | -24.5 | -23.5 | -22.5 | -23.8 | -25.2 |

**Table 4.13 Mean free energies of binding of *Agam*OBP1-, *Aaeg*OBP1- and *Cqui*OBP1-DEET obtained from two separate MD simulations**

The MD simulations with the lowest free energies of binding and the clusters of the lowest free energies derived from each simulation were subjected to free energy decomposition. The results shown in Table 4.14 predict that *Agam*OBP1 has higher binding affinity to DEET than either *Aaeg*OBP1 or *Cqui*OBP1. There is no significant difference in the three complexes with regard to van der Waals interactions, and non-polar contributions to the free energies of binding. There is a significant difference between the polar contributions to the solvation free energies with a much higher penalty for the *Agam*OBP1-DEET complex than for the *Aaeg*OBP1- and *Cqui*OBP1-DEET complexes. This penalty however is more than counterbalanced by much more favourable coulombic interactions between

*Agam*OBP1 and DEET. The higher binding affinity of *Agam*OBP1 to DEET is thus most likely due to favourable electrostatic interactions.

| | *Agam*OBP1[a] | | *Aaeg*OBP1[a] | | *Cqui*OBP1[b] | |
|---|---|---|---|---|---|---|
| | 10ns | Cluster4 | 10ns | Cluster 4 | 10ns | Cluster 3 |
| Contrib.[a] | Δ value [b] | Δ value [b] | Δ value [b] | Δ value [b] | Δ value [b] | Δ value [b] |
| $\Delta H_{vdW}$ | -32.7 | -34.4 | -32.6 | -34.3 | -33.3 | -34.3 |
| $\Delta H_{elec}$ | -7.9 | -10.2 | 0.4 | -0.5 | 1.4 | 0.9 |
| $\Delta H_{gas}$ | -40.6 | -44.7 | -32.2 | -34.8 | -32.0 | -33.4 |
| $\Delta G_{GB}$ | 19.1 | 20.6 | 12.5 | 13.7 | 11.3 | 11.9 |
| $\Delta G_{np}$ | -3.7 | -3.8 | -3.8 | -3.9 | -3.7 | -3.7 |
| $\Delta G_{solv}$ | 15.3 | 16.8 | 8.7 | 9.8 | 7.6 | 8.2 |
| $\Delta G_{gas+solv}$ | -25.3 | -27.9 | -23.5 | -25.0 | -24.3 | -25.2 |
| $\Delta TS_{total}$ | -16.9 | | -17.8 | | -18.4 | |
| $\Delta G_{total}$ | -8.4 | | -5.7 | | -5.9 | |

**Table 4.14** **Binding free energies and free energy components of OBP-DEET complexes**

Notation with regard to superscripts [a] and [b] is given in Table 3.15 on page 138

## 4.4.3 Binding of bisabolol (BIS)

### 4.4.3.1 Properties of the molecular systems used

The properties of the molecular systems used in the MD simulations are given in Table 4.15.

| | PDB id | No.aa | No. atoms | No. counterions | No. HOH | Box (Å) |
|---|---|---|---|---|---|---|
| *Agam*OBP1[a]-BIS | 2erb | 123 | 1966 | 8 | 4105 | 69x69x69 |
| *Agam*OBP1[b]-BIS | 2erb | 123 | 1966 | 8 | 4090 | 69x69x69 |
| *Aaeg*OBP1[a]-BIS | 3k1e | 124 | 1995 | 8 | 4281 | 70x70x70 |
| *Aaeg*OBP1 [b]-BIS | 3k1e | 124 | 1995 | 8 | 4349 | 70x70x70 |
| *Cqui*OBP1[a]-BIS | 3ogn | 124 | 979 | 7 | 3929 | 68x68x68 |
| *Cqui*OBP1[b]-BIS | 3ogn | 124 | 979 | 7 | 3873 | 68x68x68 |

**Table 4.15** **Bisabolol: Properties of molecular systems used in MD simulations**

No. aa & No. atoms are the number of residues and number of atoms of the protein; No. counterions is the number of ions used to neutralise the system; No. HOH is the number of water molecules contained in the solvation box

Two independent MD simulations were performed for each OBP-BIS complex. The starting conformations of BIS relative to the receptors were taken from different poses of the lowest energy docking clusters. The α-chains of the dimeric proteins were taken as the starting conformation of the OBPs. The boxed subscripts a and b after the name of the OBPs indicate one of the two independent MD simulations.

### 4.4.3.2 Time-dependent properties

The range of RMSD of the AgamOBP1 backbone Cα atoms from the respective X-ray structures is given in Table 4.16. In all but two MD simulations (*Aaeg*OBP1$_{[b]}$, *Cqui*OBP1$_{[a]}$) the differences between the minima and the maxima of RMSDs with respect to the starting structures were less than 1.1 Å.

| Ref. structure[1] | complex | residues | avg[2] (Å) | sd[3] | min[4] (Å) | max[4] (Å) | num[5] |
|---|---|---|---|---|---|---|---|
| 2erb | *Agam*OBP1$_{[a]}$-BIS | 123 | 1.24 | 0.104 | 0.93 | 1.54 | 1000 |
| | *Agam*OBP1$_{[b]}$-BIS | 123 | 1.02 | 0.164 | 0.67 | 1.53 | 1000 |
| 3k1e | *Aaeg*OBP1$_{[a]}$-BIS | 124 | 1.21 | 0.142 | 0.84 | 1.73 | 1000 |
| | *Aaeg*OBP1$_{[b]}$-BIS | 124 | 1.58 | 0.301 | 0.85 | 2.49 | 1000 |
| 3ogn | *Cqui*OBP1$_{[a]}$-BIS | 124 | 1.45 | 0.380 | 0.89 | 2.25 | 1000 |
| | *Cqui*OBP1$_{[b]}$-BIS | 124 | 1.16 | 0.197 | 0.78 | 1.74 | 1000 |

**Table 4.16** **Bisabolol: Backbone RMS deviation of MD simulation models**

1. average backbone RMSD during the trajectory with respect to the reference structure; 2. standard deviation; 3. minimum and maximum RMS deviation with respect to the reference structure; 5. number of frames

#### *Agam*OBP1-bisabolol

The time evolution of the RMSD of the Cα atoms of the protein backbone is shown in Figure 4.25 together with the interatomic distances between the hydroxyl oxygen of bisabolol (depicted as BIS@O) and the NE1 nitrogen atom of Trp112 (depicted as Trp112@NE1) and the carbonyl oxygen of Phe121 (depicted as

Phe121@O). The particular atoms were chosen as they have been shown to be involved in H-bonding and/or polar interactions in docking simulations. A similar notation (@) applies for other atomic designations. Details of H-bond formation during the MD simulations are given in Table 4.17.

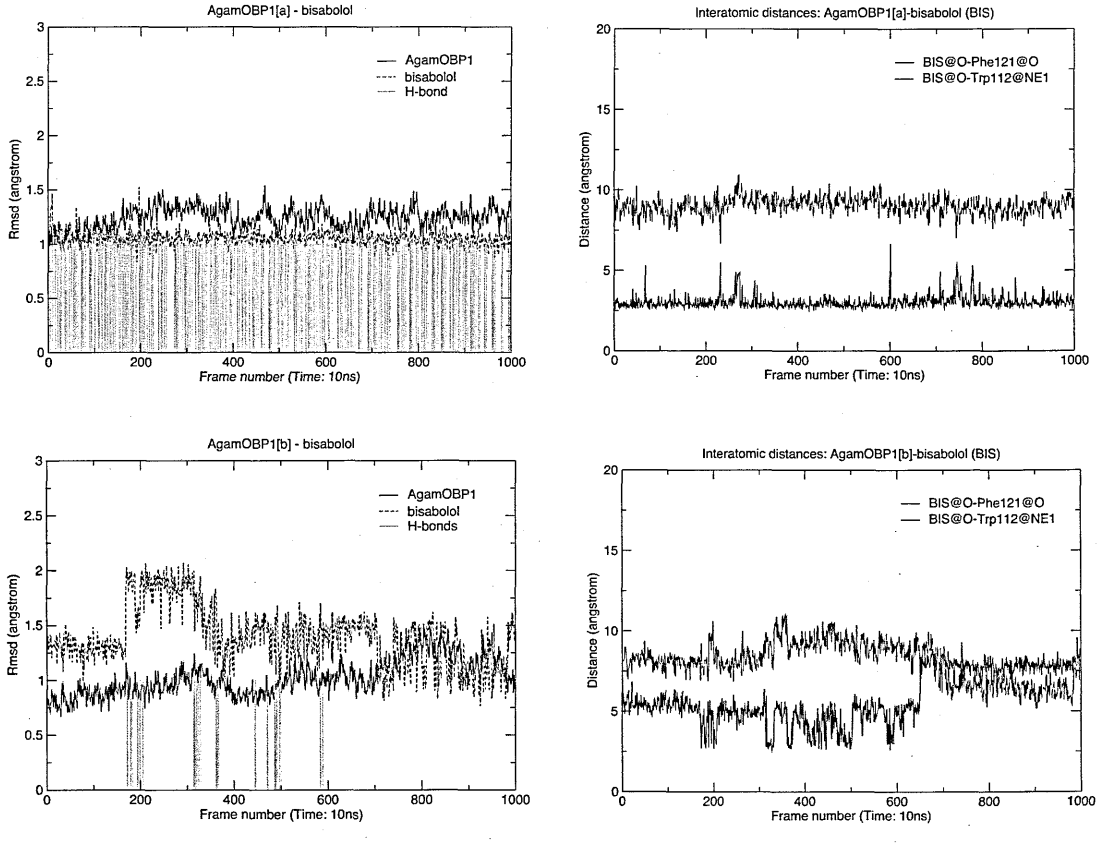



**Figure 4.25** **RMSD time series of *Agam*OBP1 Cα atoms and bisabolol from the starting structures**

| | VMD[1] | | | cpptraj[1] | | |
|---|---|---|---|---|---|---|
| | **donor** | **acceptor** | **occupancy (%)** | **donor** | **acceptor** | **occupancy (%)** |
| *Agam*OBP1[a] | BIS@O | Phe121@O | 36.1 | BIS@H1 | Phe121@O | 62 |
| *Agam*OBP1[b] | BIS@O | Phe121@O | 2.5 | BIS@H1 | Phe121@O | 5 |
| | Phe121@NH | BIS@O | 0.2 | | | |

**Table 4.17** **H-bond formation between the *Agam*OBP1 and bisabolol**

[1] See Section 2.2.6.2 H-bond formation

In the two *Agam*OBP1-BIS simulations the interatomic distances between the hydroxyl oxygen of BIS and the carbonyl oxygen of Phe121 at the start of the MD

production time were ~3.0 Å for *Agam*OBP1$_{[a]}$ and ~5.2 Å for *Agam*OBP1$_{[b]}$. H-bond formation between the hydroxyl group of bisabolol and the carbonyl oxygen of Phe121 of high occupancy was observed throughout the *Agam*OBP1$_{[a]}$-BIS MD trajectory and during the first 6ns of the simulation time of *Agam*OBP1$_{[b]}$-BIS with much less frequency. In the latter case, the functional group of BIS changed orientation abruptly and moved to within ~5.0 Å from Trp112. The interatomic distance between the oxygen atom of bisabolol and the NE1 atom of Trp112 was ~10.0 Å. The lowest free energies of binding from the *Agam*OBP1$_{[b]}$ simulation correspond to the first half of the trajectory in which H-bond formation between the ligand and Phe121 of *Agam*OBP1 was observed. Figure 4.26 shows a superposition of two representative structures, one from cluster 1 of *Agam*OBP1$_{[a]}$-BIS (lowest free energy of binding) and a second one taken from the last 3 ns of the *Agam*OBP1$_{[b]}$-BIS trajectory (lowest energy). The figure shows that with respect to the structure of the complex at the start of the trajectory, Trp112 has flipped 90° and the cyclohexene ring of BIS has flipped by almost 180° showing, as in the case of DEET (Section 4.4.2.2, Figure 4.20), that this residue may be important in stabilising interactions within the binding cavity of the protein.

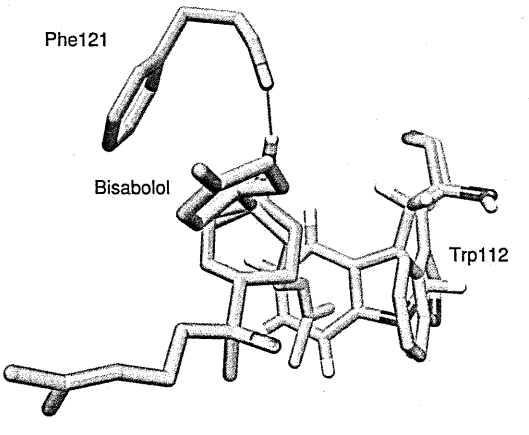



**Figure 4.26** **Bisabolol: Superposition of representative low energy structures from two separate MD simulations**

brown: *Agam*OBP1$_{[a]}$-BIS; purple: *Agam*OBP1$_{[b]}$-BIS

### *Aaeg*OBP1-bisabolol

The time evolution of the RMSD of the Cα atoms of the protein backbone is shown in Figure 4.27 together with the interatomic distances between BIS@O and Trp113@NE1. Table 4.18 gives the details of H-bond formed during the trajectory.

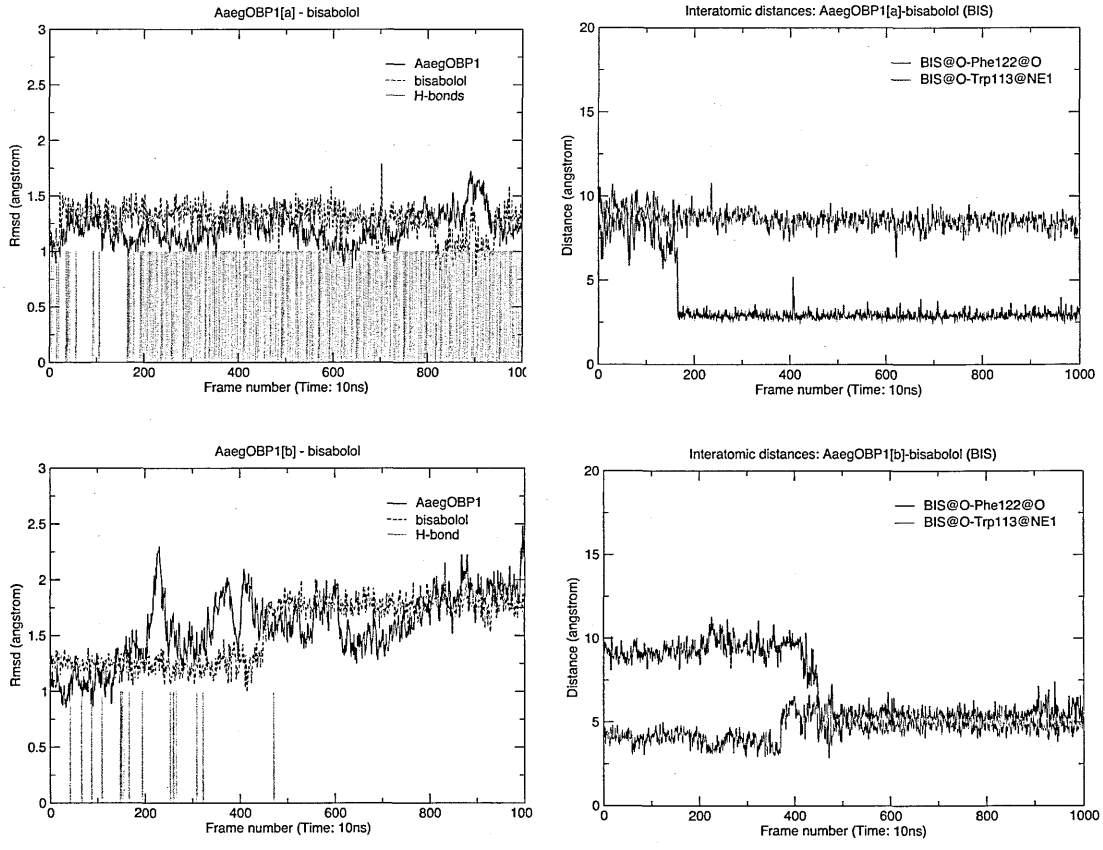



**Figure 4.27** **RMSD time series of *Aaeg*OBP1 Cα atoms and bisabolol from the starting structures**

| | VMD[1] | | | cpptraj[1] | | |
|---|---|---|---|---|---|---|
| | **donor** | **acceptor** | **occupancy (%)** | **donor** | **acceptor** | **occupancy (%)** |
| *Aaeg*OBP1[a] | BIS@O1 | LEU57@O | 1.3 | BIS@H1 | PHE@O | 59 |
| | BIS5@O1 | PHE122@O | 40.7 | BIS@H1 | LEU@O | 5 |
| | | | | HIE110@HE2 | BIS@O1 | 1 |
| *Aaeg*OBP1[b] | BIS@O1 | TRP113@CZ2 | 1.6 | BIS@H1 | ILE@O | < 0.1 |
| | BIS@O1 | TRP113@CE2 | 0.1 | BIS@H1 | TRP@NE1 | < 0.1 |
| | BIS@O1 | PHE122@O | 0.1 | BIS@H1 | PHE@O | < 0.1 |

**Table 4.18** **H-bond formation between the *Aaeg*OBP1 and bisabolol**

[1] See Section 2.2.6.2 H-bond formation

In the *Aaeg*OBP1$_{[a]}$-BIS MD simulation, at the start of the trajectory, the oxygen atom of the hydroxyl group of bisabolol was almost equidistant from the carbonyl oxygen of Phe122 and the NE1 nitrogen atom of Trp113 (~10.0 Å). During the first 2ns of the simulation time, bisabolol was seen to form H-bonds with peptidic oxygen of Leu57. Approximately 2ns from the start of the simulation time, bisabolol underwent an abrupt conformational transition with its hydroxyl group coming within H-bond distance from the carbonyl oxygen of Phe122. Henceforth and for the duration of the MD simulation, H-bonding between bisabolol and AaegOBP1 was maintained.

The second MD simulation (*Aaeg*OBP1$_{[b]}$-BIS) showed a similar pattern in which an abrupt conformational transition of bisabolol was observed at ~5ns from the start of the production MD as evidenced by a sharp change in the RMSD of the ligand and a reduction of the interatomic distance between the hydroxyl oxygen of bisabolol and peptidic oxygen atom of Phe122 bringing the two atoms to within 4-5 Å separation.

### *Cqui*OBP1-bisabolol

The time evolution of the RMSD of the Cα atoms of the protein backbone is shown in Figure 4.28 together with the interatomic distances between BIS@O and Trp113@NE1, as well as H-bonding interactions, which are given in detail in Table 4.19.

In the *Cqui*OBP1$_{[a]}$-BIS MD simulation, H-bond formation between Phe122 and bisabolol was observed throughout the trajectory despite a sudden jump of the RMSD of the protein Cα atoms by almost 1 Å. It is likely that this was due to RMSD fluctuations of loop regions of the protein that had no effect on H-bonding

interactions between bisabolol and the protein.

| | VMD[1] | | | cpptraj[1] | | |
|---|---|---|---|---|---|---|
| | donor | acceptor | occupancy (%) | donor | acceptor | occupancy (%) |
| *CquiOBP1[a]* | BIS@O | PHE122@O | 37.1 | BIS@OH | PHE_122@O | 74 |
| | PHE122@NH | BIS@O | 0.2 | WAT@H | BIS_125@O | 19 |
| | WAT@O | BIS@O | < 0.1 | | | |
| *CquiOBP1[b]* | WAT@O | BIS@O | 12 | WAT@H | BIS_125@O | 19 |
| | BIS@OH | TRP113@CE | 0.3 | BIS@OH | PHE_122@O | 9 |
| | BIS@OH | PHE122@O | 1.8 | PHE122@H | BIS@O | 2 |
| | PHE122@NH | BIS@O | 1.2 | | | |

**Table 4.19** **H-bond formation between the *CquiOBP1* and bisabolol**

[1] See Section 2.2.6.2 H-bond formation

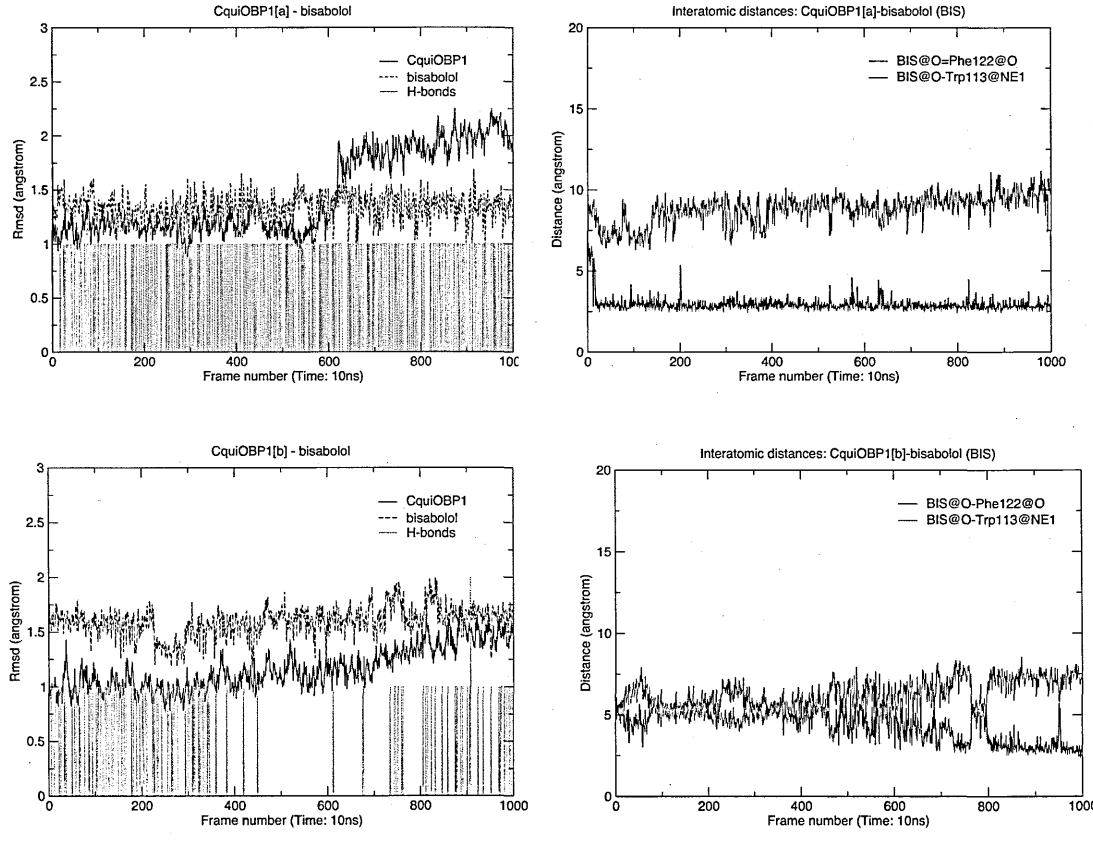



**Figure 4.28** **RMSD time series of *CquiOBP1* Cα atoms and bisabolol from the starting structures**

Another feature of the MD simulation is that the interatomic distance between the hydroxyl oxygen of bisabolol and the carbonyl oxygen of Phe122 at the start of the

MD simulation was ~ 6.5 Å dropping to within H-bonding distance shortly after (0.1ns). This is very similar to the abrupt changes in interatomic distances observed in the two *Aaeg*OBP1-BIS trajectories described in Section 4.4.3.2. In the *Cqui*OBP1$_{[b]}$-BIS MD simulation, extensive H-bond formation between bisabolol and water molecules was observed for approximately 4ns from the start of the simulation. During this part of the simulation bisabolol formed H-bonds of low occupancy with Trp113. The RMSD plot shows an upward drift of the RMSD of the Cα atoms 4ns after the start of the trajectory with a concomitant reduction of the interatomic separation between the hydroxyl oxygen of bisabolol and the carbonyl oxygen of Phe122 to within H-bond distance formation. The occupancy of H-bonds between bisabolol and Phe122 steadily increased during the second half of the simulation and in the last ~0.5ns the occupancy was > 40%.

The free energy of binding of the *Cqui*OPB1$_{[a]}$-BIS complex was ~ -31.2 kcalmol$^{-1}$ compared to -28.8 kcalmol$^{-1}$ of the *Cqui*OPB1$_{[a]}$-BIS complex. The difference was accounted for by stronger electrostatic interactions of ~3 kcalmol$^{-1}$ between the ligand and the receptor in the *Cqui*OPB1$_{[a]}$-BIS MD simulation.

### 4.4.3.3 *Time-independent properties*

#### Binding free energy components

Binding free energies for the mosquito OBPs in complex with bisabolol were calculated as described in 2.3.1.3. The free energies of binding obtained from the individual MD simulations are summarised in Table 4.20:

| | *Agam*OBP1 [a] | *Agam*OBP1 [b] | *Aaeg*OBP1 [a] | *Aaeg*OBP1 [b] | *Cqui*OBP1 [a] | *Cqui*OBP1 [b] |
|---|---|---|---|---|---|---|
| $\Delta G$ (kcal mol$^{-1}$) | -30.2 | -26.5 | -31.2 | -30.0 | -31.2 | -28.8 |

**Table 4.20** **Mean free energies of binding of *Agam*OBP1-, *Aaeg*OBP1- and *Cqui*OBP1-BIS obtained from two separate MD simulations**

The MD simulations with the lowest free energies of binding and the derivative clusters of the lowest free energies were subjected to free energy decomposition. The results are shown in Table 4.21.

| | *Agam*OBP1[a] | | *Aaeg*OBP1[a] | | *Cqui*OBP1[a] | |
|---|---|---|---|---|---|---|
| | 10ns | Cluster 1 | 10ns | Cluster 3 | 10ns | Cluster 4 |
| Contrib.[a] | Δ value [b] | Δ value [b] | Δ value [b] | Δ value [b] | Δ value [b] | Δ value [b] |
| $\Delta H_{vdW}$ | -36.2 | -36.5 | -37.8 | -37.8 | -37.4 | -38.1 |
| $\Delta H_{elec}$ | -8.2 | -8.4 | -8.5 | -9.1 | -5.3 | -4.8 |
| $\Delta H_{gas}$ | -44.4 | -44.9 | -46.2 | -46.9 | -42.7 | -42.8 |
| $\Delta G_{GB}$ | 18.4 | 18.1 | 19.4 | 19.5 | 15.9 | 15.9 |
| $\Delta G_{np}$ | -4.2 | -4.3 | -4.5 | -4.4 | -4.3 | -4.3 |
| $\Delta G_{solv}$ | 14.2 | 13.8 | 15.0 | 15.1 | 11.6 | 11.5 |
| $\Delta G_{gas+solv}$ | -30.2 | -31.1 | -31.2 | -31.8 | -31.2 | -31.3 |
| $\Delta TS_{total}$ | -17.8 | | 17.3 | | -19.4 | |
| $\Delta G_{total}$ | -12.4 | | -13.9 | | -11.8 | |

**Table 4.21** **Binding free energies and free energy components of OBP-BIS complexes**
Notation with regard to superscripts [a] and [b] is given in Table 3.15, p. 138

The data presented in Table 4.22 show that the mean free energies calculated over the duration of the 10ns trajectories do not differ significantly from those of the lowest energy clusters. Differences in $\Delta G_{gas+solv}$ of the three complexes were to within ~1 kcalmol$^{-1}$. Non-polar contributions to the solvation free energies and the van der Waals contibution to $\Delta G_{gas}$ are very similar in the three complexes. Electrostatic interactions between the OBPs and bisabolol are more favourable for *Agam*OPB1 and *Aaeg*OBP1 than for *Cqui*OBP1 but this effect is compensated by a smaller penalty of the polar contribution to the free energy of solvation in the case of *Cqui*OBP1. In view of the very small difference in calculated $\Delta G_{gas+solv}$ values and the large standard deviation in ΔS values, it is difficult to predict whether any of these OBPs shows differential binding towards bisabolol. Longer or

multiple simulations would be necessary to scan the conformational space and establish convergence to a minimum in the free energies of binding.

## 4.4.4 Binding of indole (IND)

### 4.4.4.1 Properties of the molecular systems used

The properties of the molecular systems used in the MD simulations are given in Table 4.22. The starting conformation of indole (IND) relative to *Agam*OBP1 was taken from a docking simulation in which indole was shown to form H-bonds with the peptidic group of Phe121. The α-chain of *Agam*OBP1 was taken as the starting conformation.

| | PDB id | No. aa | No. atoms | No. counterions | No. HOH | Box (Å) |
|---|---|---|---|---|---|---|
| *Agam*OBP1-IND | 2erb | 123 | 1966 | 8 | 4111 | 69x69x69 |

**Table 4.22** **Indole: Properties of molecular systems used in MD simulations**

No. aa & No. atoms are the number of residues and number of atoms of the protein; No. counterions is the number of ions used to neutralise the system; No. HOH is the number of water molecules contained in the solvation box

### 4.4.4.2 Time-dependent properties

The range of RMSD of the *Agam*OBP1 backbone Cα atoms from the respective X-ray structures is given in Table 4.24. The difference between the minima and the maxima of RMSDs with respect to the starting structures were less than 1.0 Å.

| Ref. structure[1] | complex | residues | avg[2] (Å) | sd[3] | min[4] (Å) | max[4] (Å) | num[5] |
|---|---|---|---|---|---|---|---|
| 2erb | *Agam*OBP1-IND | 123 | 1.1 | 0.14 | 0.7 | 1.7 | 1000 |

**Table 4.23** **Indole: Backbone RMS deviation of the MD simulation model**

1. reference structure; 2. average backbone RMSD during the trajectory with respect to the reference structure; 3. standard deviation; 4. minimum and maximum RMS deviation with respect to the reference structure; 5. number of frames

The time evolution of the RMSD of the Cα atoms of the protein backbone is shown in Figure 4.29 together with the interatomic distances between the nitrogen of the pyrrole ring of indole (depicted as IND@N) and the NE1 nitrogen atom of Trp112 (depicted as Trp112@NE1) and the carbonyl oxygen of Phe121 (depicted as Phe121@O). Details of H-bonding interactions are given in Table 4.25.

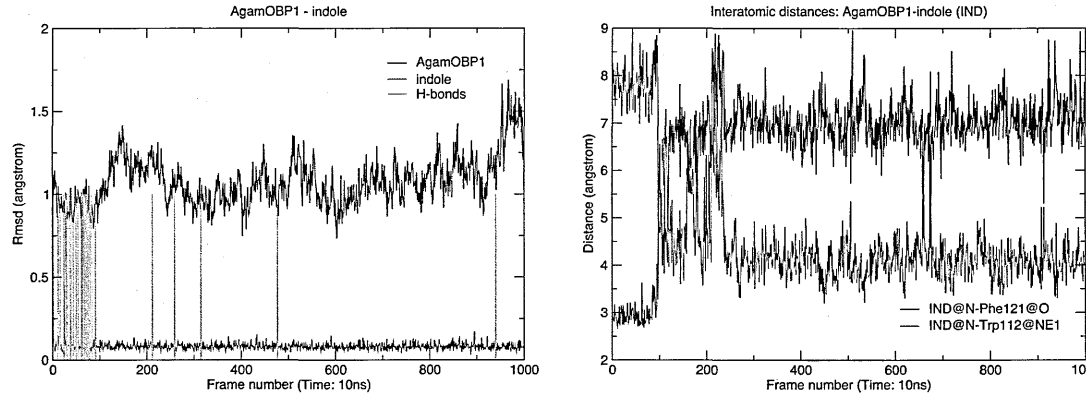



**Figure 4.29** **RMSD time series of *Agam*OBP1 Cα atoms and indole from the starting structures**

| | VMD[1] | | | cpptraj[1] | | |
|---|---|---|---|---|---|---|
| | **donor** | **acceptor** | **occupancy (%)** | **donor** | **acceptor** | **occupancy (%)** |
| *Agam*OBP1 | IND@NH | Phe121@O | 3.10 | IND@NH | Phe121@O | 8 |
| | IND@NH | Tyr8@O | 0.10 | IND@NH | Trp112@NE1 | 1 |
| | IND@NH | Trp112@CZ3 | 0.20 | | | |
| | IND@NH | Trp112@CE2 | 0.20 | | | |

**Table 4.24** **H-bond formation between the *Agam*OBP1 and indole**

[1] See Section 2.2.6.2 H-bond formation

Figure 4.29 shows indole forming H-bonds with Phe121 for the first nanoseconds of the MD simulation after which the interatomic distance between the nitrogen atom of the pyrrole ring of indole and the peptidic oxygen of Phe121 changed sharply. Poses of indole at the start and end of the simulation are shown in Figure 4.30.

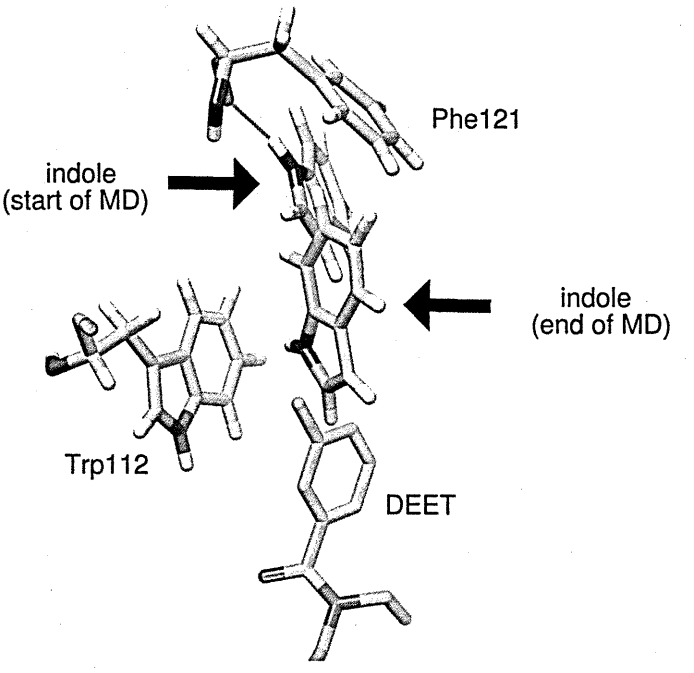
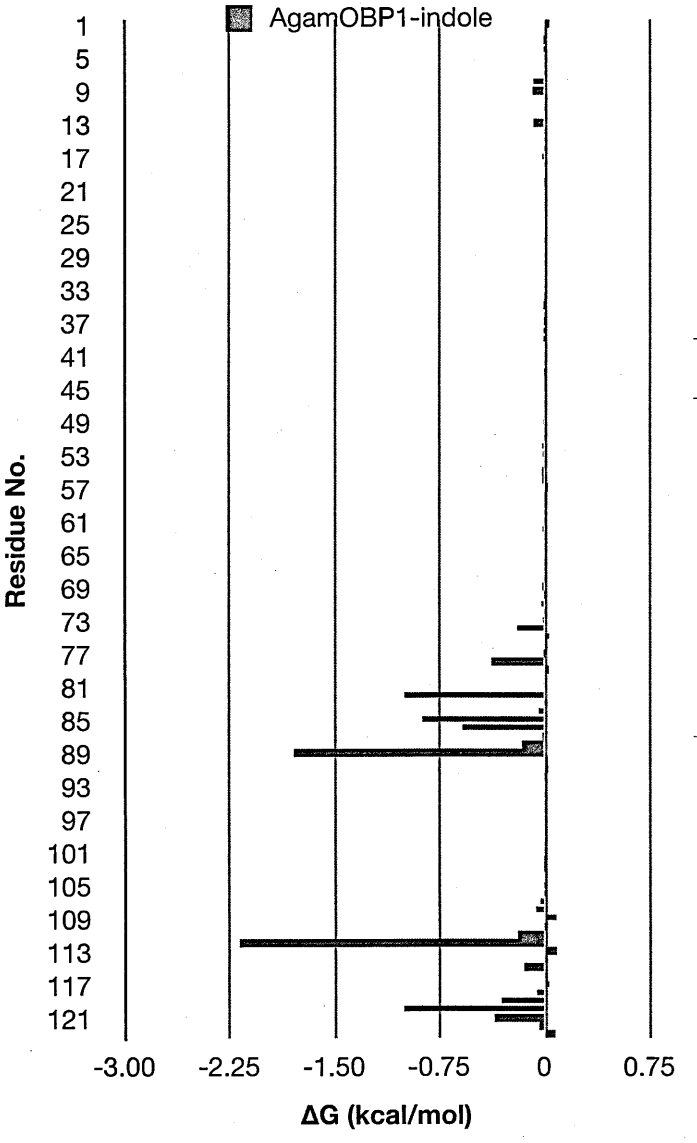



**Figure 4.30** **Relative positions of indole at the start and end of the MD *Agam*OBP1-indole simulation**

H-bond between indole and Phe121 shown in cyan

The above figure shows the H-bond formed between indole and Phe121 at the beginning of the simulation time, as well as the relative position of indole at the end of the simulation. It is recalled that the docking simulation of the *Agam*OBP1-indole (Figure 4.9) showed that the benzene rings of DEET and indole were almost coincident, suggesting that a longer simulation time may be necessary to obtain more definitive results. Per-residue decomposition of free energies shows that indole interacts mainly with residues of the 4th and 5th helices at the dimer interface (Figure 4.31).




| Residue No. | ΔG (kcalmol$^{-1}$) |
|---|---|
| Trp112 | -2.2 |
| Met89 | -1.8 |
| Tyr120 | -1.0 |
| Met82 | -1.0 |
| Ile85 | -0.9 |
| Ala86 | -0.6 |
| Leu78 | -0.4 |
| Phe121 | -0.4 |
| His119 | -0.3 |

**Figure 4.31** **Per-residue decomposition of $G_{gas+solv}$ of *Agam*OBP1-indole**

### 4.4.4.3 Time-independent properties: binding free energy components

The free energy of binding obtained from the 10ns MD production simulation are summarised in Table 4.26 and is compared to those obtained from the MD simulations with MOP, DEET and bisabolol.

| | indole | Bisabolol | DEET | MOP |
|---|---|---|---|---|
| Contrib.[a] | Δ value [b] | Δ value [b] | Δ value [b] | Δ value [b] |
| $\Delta H_{vdW}$ | -19.9 | -36.2 | -32.7 | -48.2 |
| $\Delta H_{elec}$ | -5.6 | -8.2 | -7.9 | -2.0 |
| $\Delta H_{gas}$ | -25.5 | -44.4 | -40.6 | -50.2 |
| $\Delta G_{GB}$ | 14.1 | 18.4 | 19.1 | 18.3 |
| $\Delta G_{np}$ | -2.6 | -4.2 | -3.7 | -5.9 |
| $\Delta G_{solv}$ | 11.6 | 14.2 | 15.3 | 12.5 |
| $\Delta G_{gas+solv}$ | -14.0 | -30.2 | -25.3 | -37.7 |
| $\Delta TS_{total}$ | -12.6 | -17.8 | -16.9 | -23.5 |
| $\Delta G_{total}$ | -1.4 | -12.4 | -8.4 | -14.2 |

**Table 4.25** **Binding free energies and free energy components of *Agam*OBP-ligand complexes**

Notation with regard to superscripts [a] and [b] is given in Table 3.15, p.138

The very large difference in $\Delta G_{gas+solv}$ between the *Agam*OBP1-indole and the other complexes is attributable to the much smaller contribution of hydrophobic and van der Waals interactions between indole and *Agam*OBP1.

The results presented in this Chapter suggest that there may be some subtle features that differentiate the mosquito OBPs studied in terms of odorant selectivity. Figure 4.32 shows that *Agam*OBP1-ligand binding is not favoured in the region that is proximal to the terminal "lid". Bisabolol, DEET and indole showed abrupt translational movement from positions in which the polar moieties of the molecules were within H-bond distance from the peptidic oxygen or the amide nitrogen of Phe123 to positions distal to it towards the channel of the binding cavity leading to the interface of the two protein subunits. The opposite effect, shown in Figure 4.33 was observed in the case of *Aaeg*OBP1 and, to *Cqui*OBP1. In these OBPs, ligands showed a propensity to move from positions away from the terminal "lid" towards it.

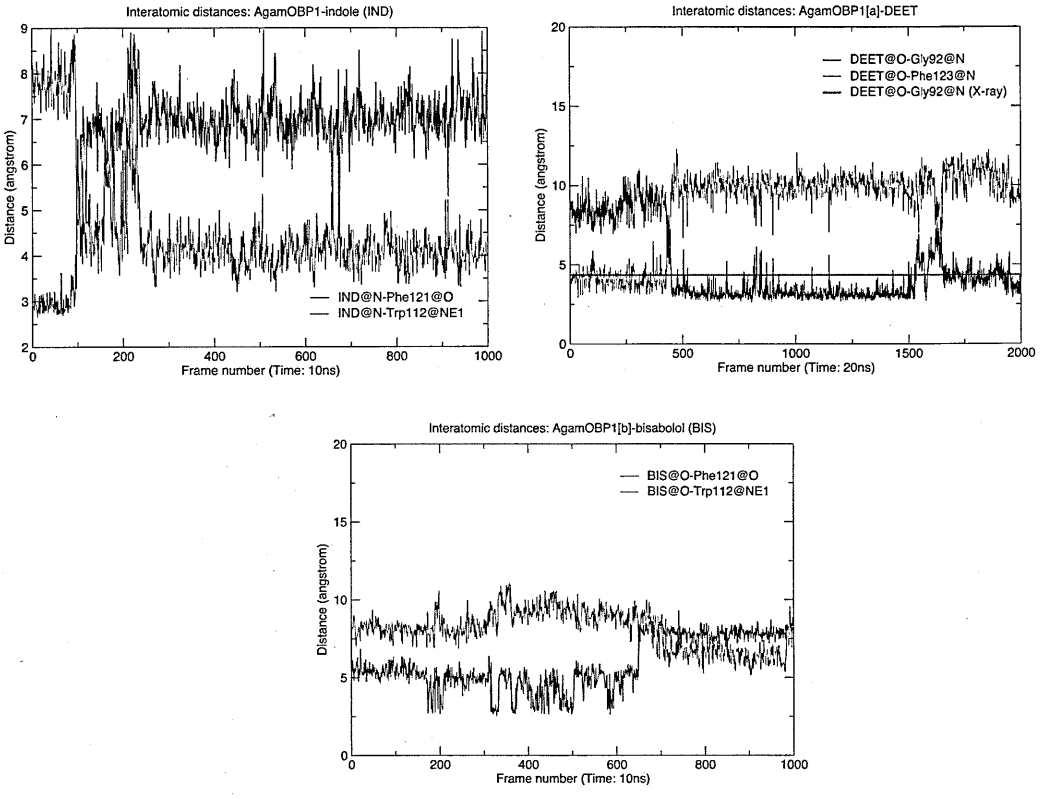



**Figure 4.32** Time evolved interatomic distances: dissociation of H-bonds

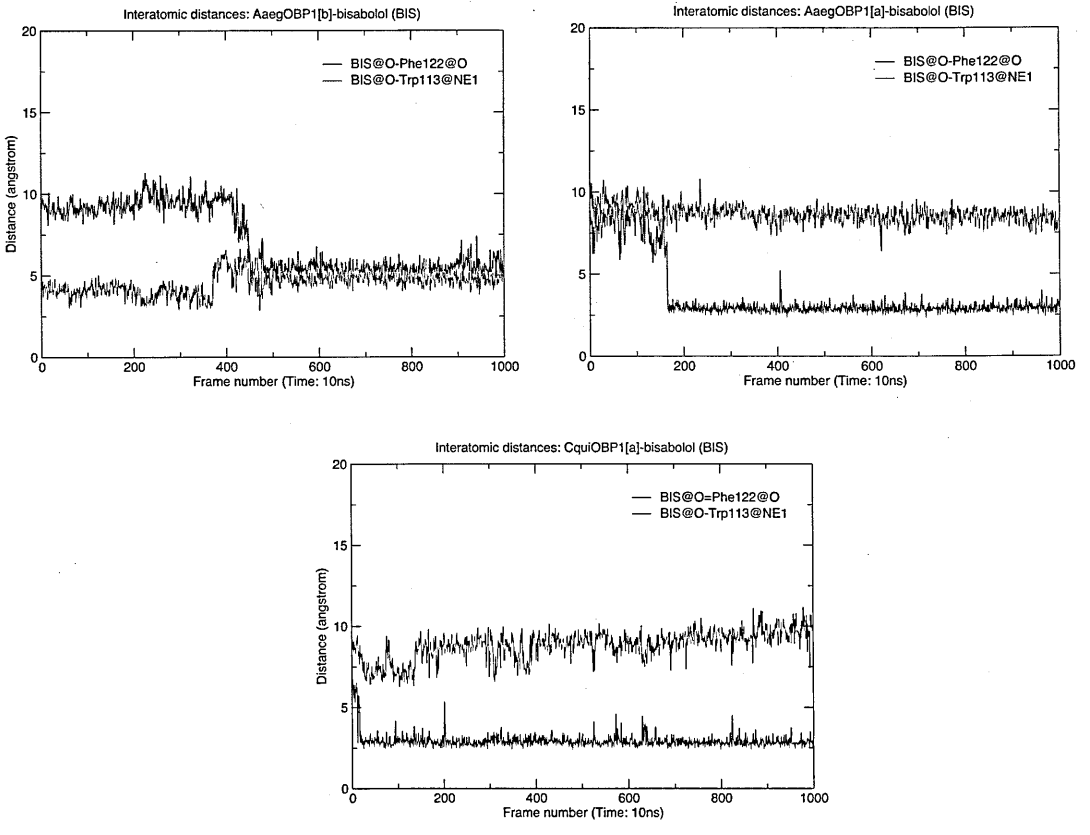



**Figure 4.33** Time evolved interatomic distances: formation of H-bonds

Similar results were obtained from the docking simulations, particularly those involving compounds with electronegative moieties such as octanoic acid as shown in Figure 4.34.

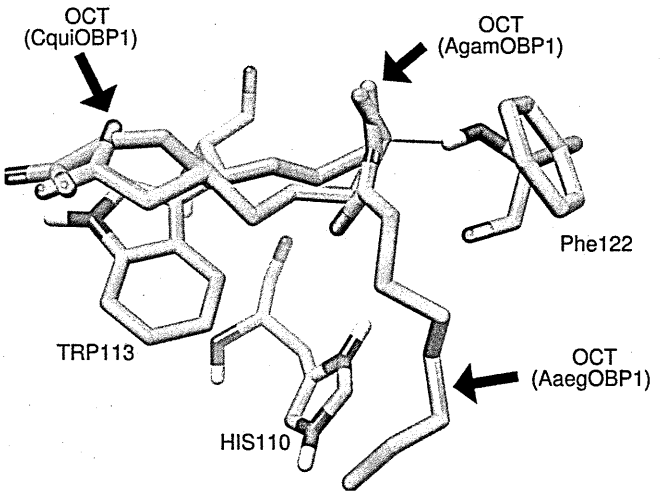
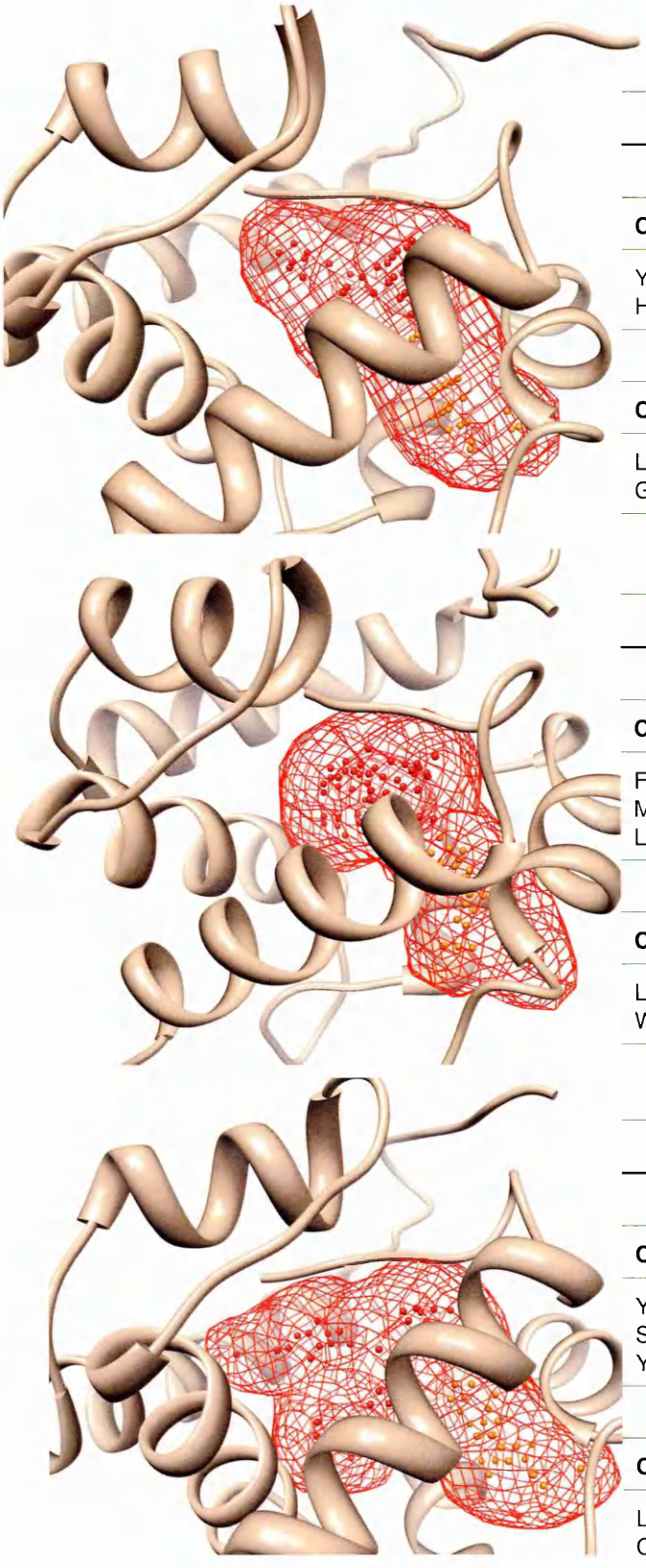



**Figure 4.34** **Docking simulations: Poses of octanoic acid (OCT) in the binding cavities of *Agam*OBP1, *Aaeg*OBP1 and *Cqui*OBP1**

The results of the docking and MD simulations suggest that ligands which can act both as a H-bond donor and acceptor such as octanoic acid and bisabolol have a propensity to form H-bonds with residues at the C-terminus of the OBPs. For example, contrary to the DEET MD trajectories and the *Agam*OBP1-DEET X-ray model, bisabolol was shown to be involved in H-bonding interactions with the receptors of high occupancy in most MD simulations.

To detect whether there are subtle differences amongst the three mosquito OBPs in terms of capacity to form H-bonds, the proteins were further characterised in terms of interactions energies as described in Sections 2.2.8 and 3.3.3. Hydroxyl (OA) probes were used to identify sub-sites favouring electrostatic interactions. The results are shown in Figure 4.35.



| *Agam*OBP1 | | |
|---|---|---|
| | **Vol.** (Å$^3$) | **TIE (kcal mol$^{-1}$)** |
| **Cluster 1** | 32 | -82.4 |
| Y10; L15; L19; L58; F59; L76; L80; M84; M91; H111; W114; H121; Y122; F123; L124; V125 | | |
| | **Vol.** (Å$^3$) | **TIE (kJmol$^{-1}$)** |
| **Cluster 2** | 32 | -71.9 |
| L73; L76; H77; L80; M84; A88; M89; M91; G92; C95; W114 | | |

| *Aaeg*OBP1 | | |
|---|---|---|
| | **Vol.** (Å$^3$) | **TIE (kcal mol$^{-1}$)** |
| **Cluster 1** | 52 | -127.5 |
| F15; M19; L58; F59; A62; V64; L76; A79; L80; M84; M91; H111; W114; H121; Y122; F123; L124; I125 | | |
| | **Vol.** (Å$^3$) | **TIE (kcal mol$^{-1}$)** |
| **Cluster 2** | 27 | -64.0 |
| L73; L76; H77; L80; A88; L89; M91; G92; K93; W114 | | |

| *Cqui*OBP1 | | |
|---|---|---|
| | **Vol.** (Å$^3$) | **TIE (kcal mol$^{-1}$)** |
| **Cluster 1** | 39 | -99.7 |
| Y10; L19; H23; Y54; L58; F59; A62; V64; L76; S79; L80; M84; A88; M91; H111; W114; H121; Y122; F123; L124; V125 | | |
| | **Vol.** (Å$^3$) | **TIE (kcal mol$^{-1}$)** |
| **Cluster 2** | 27 | -257.3 |
| L73; L76; H77; L80; A88; M89; M91; G92; K93; C95; L96; W114 | | |

**Figure 4.35** **Clusters of highest total interaction energies for the OA probe**

The figure shows clearly two contiguous clusters in all three OBPs. The cluster of the highest total interaction energy (TIE) is depicted with red spheres, whereas the

cluster with the lower TIE is depicted with orange spheres enclosed in a mesh grid. The relative location of the three clusters is the same in the three OBPs and the clusters are surrounded by the same residues with very few exceptions. The clusters of the higher TIEs border the C-terminus "lid" of the OBPs and rank in the order:

*Aaeg*OBP1 > *Cqui*OBP1 > *Agam*OBP1

with the TIE of *Aaeg*OBP1 being higher by ~15% and ~30% than those of *Cqui*OBP1 and *Agam*OBP1, respectively.

The lower energy clusters are surrounded by residues of the 4th and 5th helices. The TIEs of these latter clusters rank in the order:

*Agam*OBP1 > *Aaeg*OBP1 ≥ *Cqui*OBP1

The above results are consistent with the results of the docking and MD simulations and provide an explanation regarding the propensity of more polar ligands to form H-bonds with Phe123 of the C-terminus of *Aaeg*OBP1 and to a lesser extent with Phe123 of *Cqui*OBP1.

Experimental proof of the above prediction may have important implications as it would provide evidence of differential selectivity of these proteins. The C-terminus of the OBPs forms a "lid" over the binding pocket of the proteins. As mentioned earlier, the "lid" is thought to play the role of a gate for the release of the bound odorants. The current consensus is that the "lid" is likely to open at the acidic pH of the environment surrounding the OSNs thus releasing the odorant, which would then interact directly with the OR or co-receptor [69] [67].

In the light of the MD and docking simulations carried out in this study, it is predicted that *Aaeg*OBP1, and maybe *Cqui*OBP1, may be more selective for ligands containing electronegative moieties than *Agam*OBP1.

## *CHAPTER 5*

# *Summary and Conclusions*

The role of OBPs in olfactory function was presented in earlier chapters. Evidence to date suggests that, in some cases at least, they may be potential targets for modifying olfactory behaviour. This work has focused on the differential binding of ligands on moth and mosquito OBPs. The main approach taken was to use a combination of docking and MD simulations, the latter coupled with calculation of relative free energies of binding. Such a combined approach has been found to be more precise in the identification of the "correct" ligand binding poses than approaches based solely on docking simulations [202].

The Molecular Mechanics Poisson–Boltzmann Surface Area (MM-PBSA) method was used to calculate free energies of binding. More specifically, an implicit solvent model based on Generalised Born theory was used to perform the calculations. This, so-called MM-GBSA variant has been shown to be a computationally more efficient approximation to the Poisson theory [203]. The method has been applied in binding free energy calculations of proteins yielding insights on the stability of complexes and binding mechanisms. Furthermore, the method is appropriate for the identification of the determinants of binding. The latter can be achieved by applying a per-residue decomposition of the free energy of binding [204].

### 5.1.1 Binding affinities

Table 5.1 below summarises the effective free energies of binding ($\Delta G_{gas+solv}$) of the moth OBP/ligand complexes obtained in this study. Table 5.2 summarises effective free energies of binding of the mosquito OBP/ligand complexes. Configurational entropy results have not been been accounted for due to the inherent limitations of the method (see below). Furthermore, effective binding

energy is an adequate metric to determine the relative free energy of similar ligands.

The interpretation of the results presented in these tables needs to take into account the following caveats. Experience from applications of MM-PBSA in lead optimisation has shown a statistical uncertainty in the range of 0.24 - 1.4 kcal $mol^{-1}$ in the mean absolute deviation in the hydration free energy for neutral molecules [205]. It has also been shown that, for ligands with small binding affinity ranges, the standard deviation in the MM-PBSA free energy of binding is in the range of 0.7 - 1.0 $kcalmol^{-1}$ [206].

| | ***Bmor*PBP1** | σ | ***Msex*PBP1** | σ | ***Bmor*GOBP2** | σ | ***Msex*GOBP2** | σ |
|---|---|---|---|---|---|---|---|---|
| $\Delta G_{gas+solv}$ (BOM) | -36.6 | 2.9 | -37.8 | 2.7 | -43.7 | 2.9 | -44.9 | 2.6 |
| $\Delta G_{gas+solv}$ (BAL) | -34.6 | 2.5 | -35.5 | 2.7 | -36.0 | 2.5 | -38.8 | 2.5 |

**Table 5.1** **Summary of "effective" free energies of binding of moth OBP/ligand complexes**

σ = standard deviation

| | **Indole** | σ | **Bisabolol** | σ | **DEET** | σ | **MOP** | σ |
|---|---|---|---|---|---|---|---|---|
| $\Delta G_{gas+solv}$ (Agam) | -14.0 | 1.90 | -30.2 | 2.2 | -25.3 | 2.4 | -37.7 | 2.7 |
| $\Delta G_{gas+solv}$ (Aaeg) | NA | NA | -31.2 | 2.2 | -23.5 | 1.8 | -41.3 | 2.6 |
| $\Delta G_{gas+solv}$ (Cqui)_ | NA | NA | -31.2 | 2.0 | -24.3 | 1.8 | -40.9 | 2.6 |

**Table 5.2** **Summary of "effective" free energies of binding of mosquito OBP/ligand complexes**

σ = standard deviation; NA = not available

Taking into consideration the above, it is concluded that the affinities of *B. mori* and *M. sexta* GOBPs towards bombykol and bombykal are considerably higher than the binding affinities of the corresponding PBPs. It is noted that a difference of 1 kcal $mol^{-1}$ in free energies of binding results in a 10-fold difference in complex dissociation constants.

The data presented in Table 5.2 show little or no differences in the relative binding free energies if the standard error is taken into account with the exception possibly of the *Agam*OBP1 which seems to have considerably lower binding affinity to MOP than the other two OBPs. Considerably longer simulation times and/or multiple MD simulations from different starting conformations would be needed to increase the confidence of this prediction.

Another consideration needs to be taken into account. Binding free energy calculations are a measure of an equilibrium property. It is important to consider the results of such calculations also in a biological context. For example, in moths OBP-ligand equilibrium is established in over 30 min, whereas moth response to the odorant stimulus is observed in the millisecond time scale [78]. The implication of this may be that perireceptor interactions (ligand, OBP, OR) may not be controlled thermodynamically. Therefore, the energetically most stable protein-ligand conformation does not have to be the one found in nature, if binding were to be kinetically controlled. Evidence pointing to the OBP-ligand binding being under kinetic control has been provided by studies with a pheromone-binding protein *Lymantria dispar* (*Ldis*PBP2). These studies have shown monomer/multimer equilibria and a binding process taking place in two stages. Of these, the first stage is a rapid one in which the ligand binds to an external binding site of the protein. This stage is followed by a second slower one in which the ligand is embedded and reoriented in the internal binding pocket of *Ldis*PBP2 [78].

### 5.1.2 Ligand selectivity and conformational heterogeneity

Docking and MD simulations undertaken in the course of this work have shown that ligand binding can occur at specific "hot spots" within the binding cavities of

the OBPs considered. In the case of moth OBPs, the crystallographic models of the *BmorPBP1* and *BmorGOBP2* served as yardsticks for comparison with the results obtained in this study. The docking and MD simulations reproduced the ligand poses shown in the X-ray models and predicted additional binding "hot spots". In addition, SMD simulations predicted possible ligand entry and exit pathways into and out of *Bmor*GOBP2. Pathways for ligand entry-exit into *Bmor*GOBP2 had been speculated upon [64] but not as yet proven experimentally.

Docking and MD simulations with the mosquito OBPs reproduced with accuracy the existing crystallographic models and, similarly to the moth OBP simulations, predicted distinct "hot spots" for ligand binding.

The question that arises from the above is whether such "hot spots" are biologically relevant. Is it possible, for example, a particular ligand-binding "hot spot" in the protein binding site serves for ligand recognition and a second one for ligand scavenging? Or are they induced by conformational changes brought about upon ligand binding? Future work involving a combination of MD approaches such as umbrella sampling and thermodynamic integration could provide useful insights leading, perhaps, to the design of compounds that could potentially affect insect behaviour.

Conformational heterogeneity has been observed in several cases and the flexibility of the binding pockets of OBPs has been demonstrated through the binding of diverse compounds. However, not all of the bound ligands seem to effect a conformational change of the protein that is biologically relevant [57] [70]. An example that may be particular relevant to this study is crystal structures of the *A. mellifera* antennal pheromone-binding protein (*Amel*ASP1) in complex with the main component of the queen mandibular pheromonal mixture, 9-keto-2(E)-

decanoic acid (9-ODA) and the serendipitous ligand n-butyl-benzenesulfonamide (nBBSA). ASP1 is a member of the medium-chain structural class of OBPs and, much like *Agam-* *Aaeg-* and *Cqui*OBP1, its C-terminus is locked by double hydrogen bonds between the carboxylic end of the last residue (Ile119) and two residues namely, Lys17 and Tyr48.

In the *Amel*ASP1-nBBSA complex, the ligand was shown to be in contact with the C-terminus by an H-bond involving one of the S=O groups and the N-H group of the Ile119 and a second one involving its amide N-H and the peptidic oxygen of Phe117 (Figure 5.1 A). In the *Amel*ASP1-9-ODA complex, the -$COO^-$ moiety of the ligand makes no H-bond contacts with any of the protein residues and protrudes through the mouth of the binding cavity towards the bulk solvent, whereas the carbonyl moiety at the opposite end of the ligand forms an H-bond with the N-H moiety of the terminal Ile119 residue. The 9-ODA binding site contains a second molecule, namely glycerol the primary hydroxyl group of which establishes an H-bond with peptidic C=O of Phe117.

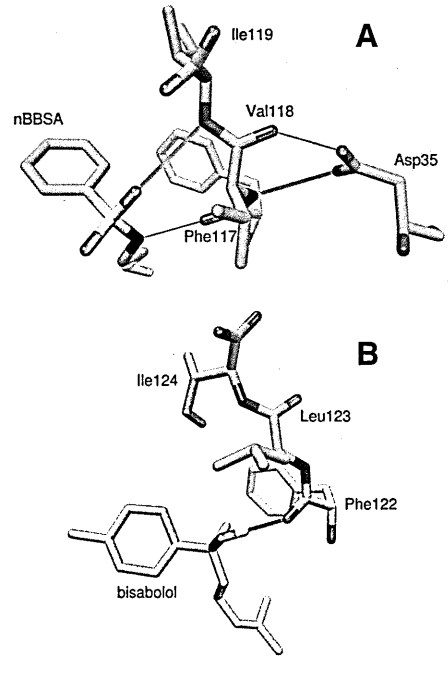



**Figure 5.1** **H-bonding interactions: *Amel*ASP1-nBBSA (A); *Aaeg*OBP1-bisabolol (B)**

This thesis reports for the first time mosquito OBP-ligand interactions that involve the formation of H-bonds between ligands and residues at the C-terminus and, in particular, Phe122. An example is shown in Figure 5.1 B which shows H-bonding interactions between bisabolol and Phe122 of *Aaeg*OBP1. This residue is conserved in the *Agam-*, *Aaeg-* and *Cqui*OBP1 and corresponds to Phe117 of *Amel*ASP1, which, as mentioned above, was shown to participate in H-bonding interactions with the ligand. The structural similarity between bisabolol and nBBSA is noticeable in the above figure. Such H-bonding interactions involving the C-terminus of the mosquito OBPs may be of biological importance by analogy to *Amel*ASP1, where it has been demonstrated that conformational plasticity at the C-terminus depends on the presence or absence of the ligand, its nature, as well as changes in pH. It has been postulated that such conformational changes may be directly linked to the binding affinity of the OBP-ligand complex to the olfactory receptor, or olfactory co-receptor, sensory neuron membrane protein (SNMP) [57].

Two of the existing mosquito OBP X-ray models have been co-crystallised with biologically relevant ligands. One of these, DEET, is bound near the interface of the two monomeric subunits away from the C-terminus "lid". The functional group of oviposition pheromone MOP, on the other hand, is in contact with residues of the C-terminus of *Cqui*OBP1 but forms no H-bonds with them. This led to the suggestion that *Cqui*OBP1 acts as a "broadband filter" that can recognise ligands on the length of their lipid chain rather than the functional group [69]. Experimental proof that the C-terminus "lid" may be also interact with specific ligands through the formation of H-bonds could have important implications in designing effective oviposition attractants for mosquito control.

### 5.1.3 The role of dimerisation

Homo- and hetero-dimerisation of OBPs/PBPs under physiological conditions has been shown to be biologically relevant (see Section 4.3.3.1). Docking simulations conducted in the course of this work showed that small ligands can bind within the main binding cavity of the mosquito OBPs as well as near or at the dimer interface with equal and possibly higher affinity. MD simulations involving OBP dimer-ligand complexes could provide extremely useful information on possible synergistic interactions, relative binding affinities and long range conformational changes. Such studies would shed some light on the potential role of the identified ligand binding "hot spots" mentioned above, which may have a role in either ligand selectivity or ligand sequestration. This is consistent with a hypothesis, which puts forward the possibility that the *Agam*OBP1 dimer may serve to bind more than one ligand in different parts of the molecule, one near the dimer interface and the other at the main binding cavity of each monomeric sub-unit. Such dual binding could be biologically relevant during insect flight through different odorant plumes [67].

The crystallographic work involving the CquiOBP1-MOP complex did not reveal any significant ligand-induced conformational change in the structure of the complex compared to those of *Agam*OBP1 and *Aaeg*OBP1 in complex with serendipitous ligands [69]. However, a striking example of extensive conformational changes upon ligand binding has been reported recently and involves the formation of a heterodimer between *Agam*OBP1 and *Agam*OBP4 following the binding of indole. It has been demonstrated that the perception of indole and 3-methyl-indole by female moths requires the presence of *Agam*OBP1 [94].

However, other evidence has shown that olfactory response to indole may require the formation of an *Agam*OBP1-*Agam*OBP4 heterodimer [200]. A recent crystallographic study revealed that *Agam*OBP4 in the presence of indole adopts a fold that is representative of the classical OBPs. However, NMR studies showed that *Agam*OBP4 in the absence of indole exists as a molten globule, which upon binding of indole undergoes a dramatic change in conformation to a well ordered structure with indole bound at one end of the main binding cavity and the concomitant formation of a second binding site for *Agam*OBP1. Although, *Agam*OBP1 is much more ordered in the absence of indole at neutral pH, conformational changes in this protein were not ruled out and they may be a requirement for the formation of the heterodimeric interface. This study provided evidence that conformational ordering plays a role in the formation of OBP heterodimers and can alter the binding affinity for ligands [194].

In this work MD simulations of *Agam*OBP1-indole complex were reported. The free energy of binding derived from MM-PBSA calculations showed that, taking into account solute entropy, there is little binding affinity between the receptor and the ligand. Not withstanding the limitations of correctly computing the contribution of solute entropy to the free energy of binding (see below) future MD work on *Agam*OBP1-*Agam*OBP4 heterodimer and indole would be instructive in identifying conformational changes induced by the ligand and the interactions at the dimer interface. Such work would also provide a measure of the effect of conformational changes on the free energy of binding.

### 5.1.4 Limitations

The main limitation of this work has been the relatively small repertoire of OBP crystal structures with little structural diversity. Another shortcoming has been that native ligands are in most cases unknown. For example it is not known which is the native ligand of *Agam*OBP1 [67]. The same applies for *Aaeg*OBP1. In the case of the *B. mori* OBPs it has been demonstrated in this work and other studies [187] that *Bmor*GOBP2 has much higher binding affinity for the pheromonal components bombykol and bombykal than *Bmor*PBP1. This begs the question whether these two paralogs have evolved for different functions and as such may be able to differentially screen different compounds. Is it possible that different OBPs are required to act in unison for the detection of pheromonal blends? The reasons for the promiscuity of OBPs in ligand binding can only be speculated.

It is possible that the PBPs from different species have different affinities for a given ligand. For example, *B. mori* and *M. sexta* utilise bombykal. However, whereas bombykal in *B. mori* functions as an antagonist of the sex pheromone bombykol, in *M. sexta* it functions as the main component of the sex pheromone blend. Differential binding of the pheromone components, i.e. *B. mori* binding bombykol and *M. sexta* bombykal could be a mechanism preventing cross-species mating. Alternatively, it is possible that OBPs have a multifunctional role. Protein multifunctionality is common and proteins are known to adapt their function differently in different chemical environments (e.g. drug resistance, ability to biodegrade chemicals). Such multifunctionality has been observed in a moth *Ostrinia* sp. and was brought about by the activation of a nonfunctional desaturase gene transcript present in the pheromone gland. Rare events such as this drive the evolution of species to respond to structurally different pheromone components [207]. Thus, in principle, an innate promiscuous capacity to host a broader spectrum of odorants than odorants representing a strict set of chemical structures would be a useful mechanism to facilitate speciation of insects. In the

early stages of this work the FlyAtlas database [208] was scanned to analyse expression data of genes coding for OBPs with a view to identifying genes that are ubiquitously expressed in the antennae. The database comprises 44 Affymetrix *Drosophila* expression arrays, each of which maps the expression of 18770 transcripts, corresponding to the vast majority of known *Drosophila* genes. It contains gene expression array data from 4 independent biological replicates in specific tissues. The database contains data for the fly head but not specifically for the fly antennae. Of the ~60 OBP genes scanned, gene expression in the head is significantly enriched only in the small subset shown in Table 5.3. Most of the OBPs shown in the table have been associated with pheromone-binding function. The rest of the OBP genes show gene expression in diverse parts of the insect body suggesting that these proteins have evolved to undertake multifunctional roles.

| OBP | SwissProt | FlyBase | Abundance | Present call | Enrichment |
|---|---|---|---|---|---|
| Obp19a | Q9VR94 | FBgn0031109 | 1079 ± 135 | 4 of 4 | 17.40 |
| Obp49a | Q7JVM1 | FBgn0050052 | 1163 ± 112 | 4 of 4 | 9.40 |
| Obp56h | Q9V8Y9 | FBgn0034475 | 220 ± 19 | 4 of 4 | 29.60 |
| Obp59a | Q86BF9 | FBgn0034766 | 121 ± 10 | 4 of 4 | 15.00 |
| lush | O02372 | FBgn0020277 | 735 ± 54 | 4 of 4 | 14.20 |
| Pbprp1 | P54191 | FBgn0011279 | 1845 ± 49 | 4 of 4 | 14.00 |
| Pbprp2 | P54192 | FBgn0011280 | 6619 ± 215 | 4 of 4 | 5.30 |
| Pbprp3 | P54193 | FBgn0011281 | 2747 ± 141 | 4 of 4 | 10.10 |
| Pbprp4 | P54194 | FBgn0011282 | 453 ± 41 | 4 of 4 | 12.90 |
| Pbprp5 | P54195 | FBgn0011283 | 1449 ± 107 | 4 of 4 | 15.60 |
| OS-C | Q23971 | FBgn0010401 | 2578 ± 89 | 4 of 4 | 12.30 |
| OS-E | Q23970 | FBgn0010403 | 3127 ± 87 | 4 of 4 | 9.70 |

**Table 5.3 OBP gene expression in the head/antennae of *D. melanogaster*** [16]

[16] Gene expression results are presented in terms of the:

- SIGNAL, which is a measure of how abundant a gene's mRNA in a given tissue. Anything over 100 is considered as being abundant and anything over 1000 as remarkable;
- PRESENT call, which is indicative of how many of the 4 tissue arrays give a detectable expression;
- ENRICHMENT, which quantifies the signal in a particular tissue in relation to that obtained from the whole fly. It is thus a measure of whether gene expression is tissue specific

This study has resulted in two findings that may have important predictive significance if proven experimentally. These are (a) the presence of binding "hot-spots" in both the moth and mosquito OBPs studied, and (b) the importance of the C-terminus "lid" of the mosquito OBPs as a selectivity filter.

Apart from the more general limitations mentioned above there are some systemic shortcomings that need to be addressed. These are summarised below.

1. Docking simulations. Screening of ligand libraries and cross-docking were performed with two different variants of AutoDock, namely, AutoDock 4.2 and Vina. The former was used for cross-docking and the latter for the screening of ligand libraries due to the fact that the algorithm can use multiple processors. However, these two variants are using different scoring functions and this has led to inconsistent results. In addition, both AutoDock variants are using Gasteiger partial charge calculation methods. It has been shown that using PM6 partial charges results in improved docking accuracy in regard to complex geometry [209].

2. MM-GBSA calculations. Although the systems considered are relatively small and 10ns MD production trajectories should provide data sets of adequate size for free energy calculations of adequate precision, longer simulation times would ensure that the conformational ensembles obtained would be properly thermalised. Furthermore, calculation of vibrational entropy by normal mode analysis leads to systematic errors as it does not take into account anharmonic contributions. MM-GBSA predictions are quite sensitive to the solute dielectric constant and this property should be parametrised properly particularly with polar ligands [192]. Another shortcoming of MM-GBSA predictions is that

structural water molecules are neglected by implicit solvation models. This can be overcome by explicitly considering such water molecules as part of solute.

3. Longer MD simulations and/or simulations from different starting conformations would be needed to allow a more effective sampling of the conformational space and thus increase confidence with regards to predictions.

4. Concentration and salt effects have not been considered. These may have important implications on dimerisation and binding. Although it is not possible to account for concentration effects by means of MD simulations, it is possible to determine the effects of salt concentration. Preliminary MM-GBSA calculations have been conducted but were inconclusive and not reported in this work.

### 5.1.5 Future directions

Any future work will have to address, at least, some of the above systemic limitations. Suggestions for additional work have been provided above. As priorities, these include:

1. MD simulations of homo- and heterodimeric OBPs with a view of detecting short- and long-range interactions upon ligand binding.
2. Applications of methods such as umbrella sampling and thermodynamic integration to decipher possible functions related to ligand-binding "hot spots".
3. Additional work to ascertain the function of the C-terminus "lid" of the medium-chain mosquito OBPs.
4. A combination of docking and MD simulations for structure-based design of compound with a potential to modify insect behaviour.

# Bibliography


1. Leal, W.S., *Pheromone Reception*, in *The Chemistry of Pheromones and Other Semiochemicals II.* 2005, Springer-Verlag Berlin Heidelberg. p. 1-36.
2. Couto, A., M. Alenius, and B.J. Dickson, *Molecular, anatomical, and functional organization of the Drosophila olfactory system.* Curr Biol, 2005. **15**(17): p. 1535-1547.
3. Franco, M.D., J. Bohbot, K. Fernandez, J. Hanna, J. Poppy, and R. Vogt, *Sensory cell proliferation within the olfactory epithelium of developing adult Manduca sexta (Lepidoptera).* PLoS One, 2007. **2**(2): p. e215.
4. Keene, A.C. and S. Waddell, *Drosophila olfactory memory: single genes to complex neural circuits.* Nat Rev Neurosci, 2007. **8**(5): p. 341-354.
5. Pelosi, P., *The role of perireceptor events in vertebrate olfaction.* Cell Mol Life Sci, 2001. **58**(4): p. 503-509.
6. Leal, W.S., *Pheromone Reception.* Top Curr Chem, 2005. **240**: p. 1-36.
7. Vogt, R.G., M.E. Rogers, M.D. Franco, and M. Sun, *A comparative study of odorant binding protein genes: differential expression of the PBP1-GOBP2 gene cluster in Manduca sexta (Lepidoptera) and the organization of OBP genes in Drosophila melanogaster (Diptera).* J Exp Biol, 2002. **205**(Pt 6): p. 719-744.
8. Picimbon, J.F. and C. Gadenne, *Evolution of noctuid pheromone binding proteins: identification of PBP in the black cutworm moth, Agrotis ipsilon.* Insect Biochem Mol Biol, 2002. **32**(8): p. 839-846.
9. Krieger, J., E. von Nickisch-Rosenegk, M. Mameli, P. Pelosi, and H. Breer, *Binding proteins from the antennae of Bombyx mori.* Insect Biochem Mol Biol, 1996. **26**(3): p. 297-307.
10. Mitsuno, H., T. Sakurai, M. Murai, T. Yasuda, S. Kugimiya, R. Ozawa, H. Toyohara, J. Takabayashi, H. Miyoshi, and T. Nishioka, *Identification of receptors of main sex-pheromone components of three Lepidopteran species.* Eur J Neurosci, 2008. **28**(5): p. 893-902.
11. Lu, T., Y.T. Qiu, G. Wang, J.Y. Kwon, M. Rutzler, H.W. Kwon, R.J. Pitts, J.J. van Loon, W. Takken, J.R. Carlson, and L.J. Zwiebel, *Odor coding in the maxillary palp of the malaria vector mosquito Anopheles gambiae.* Curr Biol, 2007. **17**(18): p. 1533-1544.
12. Pelosi, P., *Perireceptor events in olfaction.* J Neurobiol, 1996. **30**(1): p. 3-19.
13. Krieger, J. and H. Breer, *Olfactory reception in invertebrates.* Science, 1999. **286**(5440): p. 720-723.
14. Deyu, Z. and W.S. Leal, *Conformational isomers of insect odorant-binding proteins.* Arch Biochem Biophys, 2002. **397**(1): p. 99-105.
15. Kaissling, K.E., *Chemo-electrical Transduction of Insect Olfactory Receptors.* Annu Rev Neurosci, 1986. **9**: p. 121-145.
16. Honson, N., M.A. Johnson, J.E. Oliver, G.D. Prestwich, and E. Plettner, *Structure-activity studies with pheromone-binding proteins of the gypsy moth, Lymantria dispar.* Chem Senses, 2003. **28**(6): p. 479-489.
17. Xu, P., R. Atkinson, D.N. Jones, and D.P. Smith, *Drosophila OBP LUSH is required for activity of pheromone-sensitive neurons.* Neuron, 2005. **45**(2): p. 193-200.
18. Laughlin, J.D., T.S. Ha, D.N. Jones, and D.P. Smith, *Activation of pheromone-sensitive neurons is mediated by conformational activation of pheromone-binding protein.* Cell, 2008. **133**(7): p. 1255-1265.

19. Syed, Z., Y. Ishida, K. Taylor, D.A. Kimbrell, and W.S. Leal, *Pheromone reception in fruit flies expressing a moth's odorant receptor.* Proc Natl Acad Sci U S A, 2006. **103**(44): p. 16538-16543.
20. Danty, E., L. Briand, C. Michard-Vanhee, V. Perez, G. Arnold, O. Gaudemer, D. Huet, J.C. Huet, C. Ouali, C. Masson, and J.C. Pernollet, *Cloning and expression of a queen pheromone-binding protein in the honeybee: an olfactory-specific, developmentally regulated protein.* J Neurosci, 1999. **19**(17): p. 7468-7475.
21. Plettner, E., J. Lazar, E.G. Prestwich, and G.D. Prestwich, *Discrimination of pheromone enantiomers by two pheromone binding proteins from the gypsy moth Lymantria dispar.* Biochemistry, 2000. **39**(30): p. 8953-8962.
22. Pophof, B., *Moth pheromone binding proteins contribute to the excitation of olfactory receptor cells.* Naturwissenschaften, 2002. **89**(11): p. 515-518.
23. Lautenschlager, C., W.S. Leal, and J. Clardy, *Coil-to-helix transition and ligand release of Bombyx mori pheromone-binding protein.* Biochem Biophys Res Commun, 2005. **335**(4): p. 1044-1050.
24. Wetzel, C.H., H.J. Behrendt, G. Gisselmann, K.F. Stortkuhl, B. Hovemann, and H. Hatt, *Functional expression and characterization of a Drosophila odorant receptor in a heterologous cell system.* Proc Natl Acad Sci U S A, 2001. **98**(16): p. 9377-9380.
25. Vogt, R.G., *Biochemical diversity of odor detection: OBPs, ODEs and SNMPs*, in *Insect Pheromone Biochemistry and Molecular Biology*, V.R.G. Blomquist G.J, Editor. 2003, Elsevier Academic Press. p. 391-446.
26. Vogt, R.G. and L.M. Riddiford, *Pheromone binding and inactivation by moth antennae.* Nature, 1981. **293**(5828): p. 161-163.
27. Ishida, Y. and W.S. Leal, *Rapid inactivation of a moth pheromone.* Proc Natl Acad Sci U S A, 2005. **102**(39): p. 14075-14079.
28. Kaissling, K.E., *Olfactory perireceptor and receptor events in moths: a kinetic model.* Chem Senses, 2001. **26**(2): p. 125-150.
29. Vogt, R.G., L.M. Riddiford, and G.D. Prestwich, *Kinetic properties of a sex pheromone-degrading enzyme: the sensillar esterase of Antheraea polyphemus.* Proc Natl Acad Sci U S A, 1985. **82**(24): p. 8827-8831.
30. Buck, L. and R. Axel, *A novel multigene family may encode odorant receptors: A molecular basis for odor recognition.* Cell, 1991. **65**(1): p. 175-187.
31. Palczewski, K., T. Kumasaka, T. Hori, C.A. Behnke, H. Motoshima, B.A. Fox, I. Le Trong, D.C. Teller, T. Okada, R.E. Stenkamp, M. Yamamoto, and M. Miyano, *Crystal structure of rhodopsin: A G protein-coupled receptor.* Science, 2000. **289**(5480): p. 739-745.
32. Kim, J., E.N. Moriyama, C.G. Warr, P.J. Clyne, and J.R. Carlson, *Identification of novel multi-transmembrane proteins from genomic databases using quasi-periodic structural properties.* Bioinformatics, 2000. **16**(9): p. 767-775.
33. Warr, C., P. Clyne, M. de Bruyne, J. Kim, and J.R. Carlson, *Olfaction in Drosophila: Coding, Genetics and e-Genetics.* Chem Senses, 2001. **26**(2): p. 201-206.
34. Benton, R., S. Sachse, S.W. Michnick, and L.B. Vosshall, *Atypical Membrane Topology and Heteromeric Function of Drosophila Odorant Receptors In Vivo.* PLoS Biol, 2006. **4**(2): p. e20.
35. Lundin, C., L. Kall, S.A. Kreher, K. Kapp, E.L. Sonnhammer, J.R. Carlson, G. Heijne, and I. Nilsson, *Membrane topology of the Drosophila OR83b odorant receptor.* FEBS Lett, 2007. **581**(29): p. 5601-5604.

36. Smart, R., A. Kiely, M. Beale, E. Vargas, C. Carraher, A.V. Kralicek, D.L. Christie, C. Chen, R.D. Newcomb, and C.G. Warr, *Drosophila odorant receptors are novel seven transmembrane domain proteins that can signal independently of heterotrimeric G proteins.* Insect Biochem Molec Biol, 2008. **38**(8): p. 770-780.
37. Elmore, T., R. Ignell, J.R. Carlson, and D.P. Smith, *Targeted mutation of a Drosophila odor receptor defines receptor requirement in a novel class of sensillum.* J Neurosci, 2003. **23**(30): p. 9906-9912.
38. Nakagawa, T., T. Sakurai, T. Nishioka, and K. Touhara, *Insect sex-pheromone signals mediated by specific combinations of olfactory receptors.* Science, 2005. **307**(5715): p. 1638-1642.
39. Larsson, M.C., A.I. Domingos, W.D. Jones, M.E. Chiappe, H. Amrein, and L.B. Vosshall, *Or83b encodes a broadly expressed odorant receptor essential for Drosophila olfaction.* Neuron, 2004. **43**(5): p. 703-714.
40. Sato, K., M. Pellegrino, T. Nakagawa, L.B. Vosshall, and K. Touhara, *Insect olfactory receptors are heteromeric ligand-gated ion channels.* Nature, 2008. **452**(7190): p. 1002-1006.
41. Wicher, D., R. Schafer, R. Bauernfeind, M.C. Stensmyr, R. Heller, S.H. Heinemann, and B.S. Hansson, *Drosophila odorant receptors are both ligand-gated and cyclic-nucleotide-activated cation channels.* Nature, 2008. **452**(7190): p. 1007-1011.
42. Benton, R., K.S. Vannice, and L.B. Vosshall, *An essential role for a CD36-related receptor in pheromone detection in Drosophila.* Nature, 2007. **450**(7167): p. 289-293.
43. Jin, X., T.S. Ha, and D.P. Smith, *SNMP is a signaling component required for pheromone sensitivity in Drosophila.* Proc Natl Acad Sci USA, 2008. **105**(31): p. 10996-11001.
44. Croset, V., R. Rytz, S.F. Cummins, A. Budd, D. Brawand, H. Kaessmann, T.J. Gibson, and R. Benton, *Ancient protostome origin of chemosensory ionotropic glutamate receptors and the evolution of insect taste and olfaction.* PLoS Genet, 2010. **6**(8): p. e1001064.
45. Hekmat-Scafe, D.S., C.R. Scafe, A.J. McKinney, and M.A. Tanouye, *Genome-wide analysis of the odorant-binding protein gene family in Drosophila melanogaster.* Genome Res, 2002. **12**(9): p. 1357-1369.
46. Xu, P.X., L.J. Zwiebel, and D.P. Smith, *Identification of a distinct family of genes encoding atypical odorant-binding proteins in the malaria vector mosquito, Anopheles gambiae.* Insect Mol Biol, 2003. **12**(6): p. 549-560.
47. Zhou, J.J., W. Huang, G.A. Zhang, J.A. Pickett, and L.M. Field, *"Plus-C" odorant-binding protein genes in two Drosophila species and the malaria mosquito Anopheles gambiae.* Gene, 2004. **327**(1): p. 117-129.
48. Scaloni, A., M. Monti, S. Angeli, and P. Pelosi, *Structural analysis and disulfide-bridge pairing of two odorant-binding proteins from Bombyx mori.* Biochem Biophys Res Commun. **266**(2): p. 386-391.
49. Leal, W.S., L. Nikonova, and G. Peng, *Disulfide structure of the pheromone binding protein from the silkworm moth, Bombyx mori.* FEBS Lett, 1999. **464**(1-2): p. 85-90.
50. Briand L., N.C., Huet JC, Takahashi M., Pernollet JC, *Ligand binding and physico-chemical properties of ASP2, a recombinant odorant-binding protein from honeybee (Apis mellifera L.).* Eur J Biochem, 2001. **268**(3): p. 752-760.
51. Mohanty, S., S. Zubkov, and A.M. Gronenborn, *The solution NMR structure of Antheraea polyphemus PBP provides new insight into pheromone*

*recognition by pheromone-binding proteins.* J Mol Biol, 2004. **337**(2): p. 443-451.
52. Pelosi, P., J.J. Zhou, L.P. Ban, and M. Calvello, *Soluble proteins in insect chemical communication.* Cell Mol Life Sci, 2006. **63**(14): p. 1658-1676.
53. Galindo, K. and D.P. Smith, *A large family of divergent Drosophila odorant-binding proteins expressed in gustatory and olfactory sensilla.* Genetics, 2001. **159**(3): p. 1059-1072.
54. Sanchez-Gracia, A., F.G. Vieira, and J. Rozas, *Molecular evolution of the major chemosensory gene families in insects.* Heredity, 2009. **103**(3): p. 208-216.
55. Lagarde, A., S. Spinelli, M. Tegoni, X. He, L. Field, J.J. Zhou, and C. Cambillau, *The crystal structure of odorant binding protein 7 from Anopheles gambiae exhibits an outstanding adaptability of its binding site.* J Mol Biol, 2011. **414**(3): p. 401-412.
56. Lagarde, A., S. Spinelli, H. Qiao, M. Tegoni, P. Pelosi, and C. Cambillau, *Crystal structure of a novel type of odorant-binding protein from Anopheles gambiae, belonging to the C-plus class.* Biochem J, 2011. **437**(3): p. 423-430.
57. Pesenti, M.E., S. Spinelli, V. Bezirard, L. Briand, J.C. Pernollet, M. Tegoni, and C. Cambillau, *Structural basis of the honey bee PBP pheromone and pH-induced conformational change.* J Mol Biol, 2008. **380**(1): p. 158-169.
58. Sandler, B.H., L. Nikonova, W.S. Leal, and J. Clardy, *Sexual attraction in the silkworm moth: structure of the pheromone-binding-protein-bombykol complex.* Chem Biol, 2000. **7**(2): p. 143-151.
59. Horst, R., F. Damberger, P. Luginbuhl, P. Guntert, G. Peng, L. Nikonova, W.S. Leal, and K. Wuthrich, *NMR structure reveals intramolecular regulation mechanism for pheromone binding and release.* Proc Natl Acad Sci U S A, 2001. **98**(25): p. 14374-14379.
60. Keil, T.A., *Surface coats of pore tubules and olfactory sensory dendrites of a silkmoth revealed by cationic markers.* Tissue Cell, 1984. **16**(5): p. 705-717.
61. Zubkov, S., A.M. Gronenborn, I.J. Byeon, and S. Mohanty, *Structural consequences of the pH-induced conformational switch in A.polyphemus pheromone-binding protein: mechanisms of ligand release.* J Mol Biol, 2005. **354**(5): p. 1081-1090.
62. Damberger, F.F., Y. Ishida, W.S. Leal, and K. Wuthrich, *Structural basis of ligand binding and release in insect pheromone-binding proteins: NMR structure of Antheraea polyphemus PBP1 at pH 4.5.* J Mol Biol, 2007. **373**(4): p. 811-819.
63. Xu, X., W. Xu, J. Rayo, Y. Ishida, W.S. Leal, and J.B. Ames, *NMR structure of navel orangeworm moth pheromone-binding protein (AtraPBP1): implications for pH-sensitive pheromone detection.* Biochemistry, 2010. **49**(7): p. 1469-1476.
64. Zhou, J.J., G. Robertson, X. He, S. Dufour, A.M. Hooper, J.A. Pickett, N.H. Keep, and L.M. Field, *Characterisation of Bombyx mori Odorant-binding proteins reveals that a general odorant-binding protein discriminates between sex pheromone components.* J Mol Biol, 2009. **389**(3): p. 529-545.
65. Kowcun, A., N. Honson, and E. Plettner, *Olfaction in the gypsy moth, Lymantria dispar: effect of pH, ionic strength, and reductants on pheromone transport by pheromone-binding proteins.* J Biol Chem, 2001. **276**(48): p. 44770-44776.

66. Tsitsanou, K.E., T. Thireou, C.E. Drakou, K. Koussis, M.V. Keramioti, D.D. Leonidas, E. Eliopoulos, K. Iatrou, and S.E. Zographos, *Anopheles gambiae odorant binding protein crystal complex with the synthetic repellent DEET: implications for structure-based design of novel mosquito repellents.* Cell Mol Life Sci, 2012. **69**(2): p. 283-297.
67. Wogulis, M., T. Morgan, Y. Ishida, W.S. Leal, and D.K. Wilson, *The crystal structure of an odorant binding protein from Anopheles gambiae: evidence for a common ligand release mechanism.* Biochem Biophys Res Commun, 2006. **339**(1): p. 157-164.
68. Leite, N.R., R. Krogh, W. Xu, Y. Ishida, J. Iulek, W.S. Leal, and G. Oliva, *Structure of an odorant-binding protein from the mosquito Aedes aegypti suggests a binding pocket covered by a pH-sensitive "Lid".* PLoS One, 2009. **4**(11): p. e8006.
69. Mao, Y., X. Xu, W. Xu, Y. Ishida, W.S. Leal, J.B. Ames, and J. Clardy, *Crystal and solution structures of an odorant-binding protein from the southern house mosquito complexed with an oviposition pheromone.* Proc Natl Acad Sci U S A, 2010. **107**(44): p. 19102-19107.
70. Lautenschlager, C., W.S. Leal, and J. Clardy, *Bombyx mori pheromone-binding protein binding nonpheromone ligands: implications for pheromone recognition.* Structure, 2007. **15**(9): p. 1148-1154.
71. Kruse, S.W., R. Zhao, D.P. Smith, and D.N. Jones, *Structure of a specific alcohol-binding site defined by the odorant binding protein LUSH from Drosophila melanogaster.* Nat Struct Biol, 2003. **10**(9): p. 694-700.
72. Thode, A.B., S.W. Kruse, J.C. Nix, and D.N. Jones, *The role of multiple hydrogen-bonding groups in specific alcohol binding sites in proteins: insights from structural studies of LUSH.* J Mol Biol, 2008. **376**(5): p. 1360-1376.
73. Lartigue, A., A. Gruez, S. Spinelli, S. Riviere, R. Brossut, M. Tegoni, and C. Cambillau, *The crystal structure of a cockroach pheromone-binding protein suggests a new ligand binding and release mechanism.* J Biol Chem, 2003. **278**(32): p. 30213-30218.
74. Campanacci, V., S. Longhi, P. Nagnan-Le Meillour, C. Cambillau, and M. Tegoni, *Recombinant pheromone binding protein 1 from Mamestra brassicae (MbraPBP1). Functional and structural characterization.* Eur J Biochem, 1999. **264**(3): p. 707-716.
75. Wojtasek, H. and W.S. Leal, *Conformational change in the pheromone-binding protein from Bombyx mori induced by pH and by interaction with membranes.* J Biol Chem, 1999. **274**(43): p. 30950-30956.
76. Maida, R., T. Proebstl, and M. Laue, *Heterogeneity of odorant-binding proteins in the antennae of Bombyx mori.* Chem Senses, 1997. **22**(5): p. 503-515.
77. Leal, W.S., *Duality monomer-dimer of the pheromone-binding protein from Bombyx mori.* Biochem Biophys Res Commun, 2000. **268**(2): p. 521-529.
78. Gong, Y., T.C. Pace, C. Castillo, C. Bohne, M.A. O'Neill, and E. Plettner, *Ligand-interaction kinetics of the pheromone- binding protein from the gypsy moth, L. dispar: insights into the mechanism of binding and release.* Chem Biol, 2009. **16**(2): p. 162-172.
79. Tegoni, M., V. Campanacci, and C. Cambillau, *Structural aspects of sexual attraction and chemical communication in insects.* Trends Biochem Sci, 2004. **29**(5): p. 257-264.
80. Andronopoulou, E., V. Labropoulou, V. Douris, D.F. Woods, H. Biessmann, and K. Iatrou, *Specific interactions among odorant-binding proteins of the*

*African malaria vector Anopheles gambiae.* Insect Mol Biol, 2006. **15**(6): p. 797-811.
81. Feng, L. and G.D. Prestwich, *Expression and characterization of a lepidopteran general odorant binding protein.* Insect Biochem Molec Biol, 1997. **27**(5): p. 405-412.
82. Campanacci, V., J. Krieger, S. Bette, J.N. Sturgis, A. Lartigue, C. Cambillau, H. Breer, and M. Tegoni, *Revisiting the specificity of Mamestra brassicae and Antheraea polyphemus pheromone-binding proteins with a fluorescence binding assay.* J Biol Chem, 2001. **276**(23): p. 20078-20084.
83. Bette, S., H. Breer, and J. Krieger, *Probing a pheromone binding protein of the silkmoth Antheraea polyphemus by endogenous tryptophan fluorescence.* Insect Biochem Mol Biol, 2002. **32**(3): p. 241-246.
84. Shanbhag, S.R., D. Hekmat-Scafe, M.S. Kim, S.K. Park, J.R. Carlson, C. Pikielny, D.P. Smith, and R.A. Steinbrecht, *Expression mosaic of odorant-binding proteins in Drosophila olfactory organs.* Microsc Res Tech, 2001. **55**(5): p. 297-306.
85. Leal, W.S., A.M. Chen, Y. Ishida, V.P. Chiang, M.L. Erickson, T.I. Morgan, and J.M. Tsuruda, *Kinetics and molecular properties of pheromone binding and release.* Proc Natl Acad Sci U S A, 2005. **102**(15): p. 5386-5391.
86. Briand L Fau - Nespoulous, C., J.C. Nespoulous C Fau - Huet, M. Huet Jc Fau - Takahashi, J.C. Takahashi M Fau - Pernollet, and P. JC, *Ligand binding and physico-chemical properties of ASP2, a recombinant odorant-binding protein from honeybee (Apis mellifera L.).* Eur J Biochem, 2001. **268**(3): p. 752-760.
87. Riviere, S., A. Lartigue, B. Quennedey, V. Campanacci, J.P. Farine, M. Tegoni, C. Cambillau, and R. Brossut, *A pheromone-binding protein from the cockroach Leucophaea maderae: cloning, expression and pheromone binding.* Biochem J, 2003. **371**(Pt 2): p. 573-579.
88. Hooper, A.M., S. Dufour, X. He, A. Muck, J.J. Zhou, R. Almeida, L.M. Field, A. Svatos, and J.A. Pickett, *High-throughput ESI-MS analysis of binding between the Bombyx mori pheromone-binding protein BmorPBP1, its pheromone components and some analogues.* Chem Commun (Camb), 2009(38): p. 5725-5727.
89. Qiao, H., E. Tuccori, X. He, A. Gazzano, L. Field, J.J. Zhou, and P. Pelosi, *Discrimination of alarm pheromone (E)-beta-farnesene by aphid odorant-binding proteins.* Insect Biochem Mol Biol, 2009. **39**(5-6): p. 414-419.
90. Leal, W.S., A.M. Chen, and M.L. Erickson, *Selective and pH-dependent binding of a moth pheromone to a pheromone-binding protein.* J Chem Ecol, 2005. **31**(10): p. 2493-2499.
91. Maida, R., G. Ziegelberger, and K.E. Kaissling, *Ligand binding to six recombinant pheromone-binding proteins of Antheraea polyphemus and Antheraea pernyi.* J Comp Physiol B, 2003. **173**(7): p. 565-573.
92. Pophof, B., *Inhibitors of sensillar esterase block reversibly the responses of moth pheromone receptor cells.* Ann N Y Acad Sci, 1998. **855**: p. 316-319.
93. Zhou, J.J., *Odorant-binding proteins in insects.* Vitam Horm, 2010. **83**: p. 241-272.
94. Biessmann, H., E. Andronopoulou, M.R. Biessmann, V. Douris, S.D. Dimitratos, E. Eliopoulos, P.M. Guerin, K. Iatrou, R.W. Justice, T. Krober, O. Marinotti, P. Tsitoura, D.F. Woods, and M.F. Walter, *The Anopheles gambiae odorant binding protein 1 (AgamOBP1) mediates indole recognition in the antennae of female mosquitoes.* PLoS One, 2010. **5**(3): p. e9471.

95. Qiao, H., X. He, D. Schymura, L. Ban, L. Field, F. Dani, E. Michelucci, B. Caputo, A. Torre, K. Iatrou, J.-J. Zhou, J. Krieger, and P. Pelosi, *Cooperative interactions between odorant-binding proteins of Anopheles gambiae.* Cell Mol Life Sci, 2011. **68**(10): p. 1799-1813.
96. Jiang, Q.Y., W.X. Wang, Z. Zhang, and L. Zhang, *Binding specificity of locust odorant binding protein and its key binding site for initial recognition of alcohols.* Insect Biochem Mol Biol, 2009. **39**(7): p. 440-447.
97. Charlier, L., S. Antonczak, E. Jacquin-Joly, D. Cabrol-Bass, and J. Golebiowski, *Deciphering the selectivity of Bombyx mori pheromone binding protein for bombykol over bombykal: a theoretical approach.* Chemphyschem, 2008. **9**(18): p. 2785-2793.
98. Grater, F., W. Xu, W. Leal, and H. Grubmuller, *Pheromone discrimination by the pheromone-binding protein of Bombyx mori.* Structure, 2006. **14**(10): p. 1577-1586.
99. Kim, J.T., A.D. Hamilton, C.M. Bailey, R.A. Domaoal, L. Wang, K.S. Anderson, and W.L. Jorgensen, *FEP-guided selection of bicyclic heterocycles in lead optimization for non-nucleoside inhibitors of HIV-1 reverse transcriptase.* J Am Chem Soc, 2006. **128**(48): p. 15372-15373.
100. Kollman, P., *Free energy calculations: Applications to chemical and biochemical phenomena.* Chem Rev, 1993. **93**(7): p. 2395-2417.
101. Homeyer, N. and H. Gohlke, *Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method.* Mol Inf, 2012. **31**: p. 114-122.
102. Schlick, T., *Molecular Modeling and Simulation: An Interdisciplinary Guide.* Interdisciplinary Applied Mathematics, ed. J.E. Marsden, Wiggins, S., Sirovich, L., Antman, S.S. Vol. 21. 2002, New York: Springer-Verlag.
103. Leach, A.R., *Molecular Modelling: Principles and Applications.* 2nd ed. 2001, Pearson Prentice Hall, Harlow, England.
104. Adcock, S.A. and J.A. McCammon, *Molecular dynamics: survey of methods for simulating the activity of proteins.* Chem Rev, 2006. **106**(5): p. 1589-1615.
105. Leach, A.R. *Ligand-Based Approaches: Core Molecular Modeling.* Comprehensive Medicinal Chemistry II, 2007. **4**, 88-114
106. Adler, B.J., Wainwright, T.E., *Phase Transition for a Hard Sphere.* J Chem Phys, 1957. **27**(5) p. 1208-1209.
107. MacCammon, J.A., Gelin, B.R., Karplus, M., *Dynamics of folded proteins.* Nature, 1977. **267**: p. 585-590.
108. Dror, R.O., R.M. Dirks, J.P. Grossman, H. Xu, and D.E. Shaw, *Biomolecular simulation: a computational microscope for molecular biology.* Annu Rev Biophys, 2012. **41**: p. 429-452.
109. Ponder, J.W. and D.A. Case, *Force fields for protein simulations.* Adv Protein Chem, 2003. **66**: p. 27-85.
110. Allinger, N.L., K.H. Chen, J.H. Lii, and K.A. Durkin, *Alcohols, ethers, carbohydrates, and related compounds. I. The MM4 force field for simple compounds.* J Comput Chem, 2003. **24**(12): p. 1447-1472.
111. Phillips, J.C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R.D. Skeel, L. Kale, and K. Schulten, *Scalable molecular dynamics with NAMD.* J Comput Chem, 2005. **26**(16): p. 1781-1802.
112. Oostenbrink, C., A. Villa, A.E. Mark, and W.F. Van Gunsteren, *A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6.* J Comput Chem, 2004. **25**(13): p. 1656-1676.

113. Vanommeslaeghe K, H., E. Acharya, C., Kundu, S., Zhong, S., Shim, J., Darian, E., Guvench, O., Lopes, P., Vorobyov, I., Mackerell, A. D., Jr., *CHARMM general force field: A force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields.* J Comput Chem, 2010. **31**(4): p. 671-690.
114. Jorgensen, W.L. and J. Tirado-Rives, *The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin.* J Am Chem Soc, 1988. **110**(6): p. 1657-1666.
115. Darden, T., D. York, and L. Pedersen, *Particle mesh Ewald: An N·log(N) method for Ewald sums in large systems.* J Chem Phys, 1993. **98**(12): p. 10089.
116. Darden, T., L. Perera, L. Li, and L. Pedersen, *New tricks for modelers from the crystallography toolkit: the particle mesh Ewald algorithm and its use in nucleic acid simulations.* Structure, 1999. **7**(3): p. R55-60.
117. Hockney, R.W., *The potential calculation and some applications.* Meth Comput Phys, 1970. **9**: p. 135-211.
118. Verlet, L., *Computer Experiments on Classical Fluids. I. Thermodynamical Properties of Lennard-Jones Molecules.* Phys Rev, 1967. **159**: p. 98-103.
119. Swope, W.C., Andersen, H.C., Berens, P.H., Wilson, K.R., *A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters.* J Chem Phys, 1982. **76**: p. 637-649.
120. Beeman, D., *Some multistep methods for use in molecular dynamics calculations.* J Comput Phys, 1976. **20**(2): p. 130-139.
121. Gear, C.W., *Numerical Initial Value Problems in Ordinary Differential Equations.* 1971, Englewood Cliffs, NJ: Prentice-Hall.
122. Davidchack, R.L., *Discretization errors in molecular dynamics simulations with deterministic and stochastic thermostats.* J Comput Phys, 2010. **229**(24): p. 9323-9346.
123. Joung, I.S. and T.E. Cheatham, 3rd, *Determination of alkali and halide monovalent ion parameters for use in explicitly solvated biomolecular simulations.* J Phys Chem B, 2008. **112**(30): p. 9020-9041.
124. Guillot, B., *A reappraisal of what we have learnt during three decades of computer simulations on water.* J Mol Liq, 2002. **101**(1–3): p. 219-260.
125. Jorgensen, W.L.C., J.; Madura, J.D.; Impey, R.W; Klein, M.L., *Comparison of simple potential functions for simulating liquid water.* J. Chem. Phys, 1983. **79**: p. 926-935.
126. Berendsen, H.J.C., J.R. Grigera, and T.P. Straatsma, *The missing term in effective pair potentials.* J Phys Chem, 1987. **91**(24): p. 6269-6271.
127. Horn, H.W., W.C. Swope, J.W. Pitera, J.D. Madura, T.J. Dick, G.L. Hura, and T. Head-Gordon, *Development of an improved four-site water model for biomolecular simulations: TIP4P-Ew.* J Chem Phys, 2004. **120**(20): p. 9665-9678.
128. Baker, N.A., *Improving implicit solvent simulations: a Poisson-centric view.* Curr Opin Struct Biol, 2005. **15**(2): p. 137-143.
129. Roux, B.t. and T. Simonson, *Implicit solvent models.* Biophys Chem, 1999. **78**(1–2): p. 1-20.
130. Dong, F., B. Olsen, and N.A. Baker, *Computational methods for biomolecular electrostatics.* Methods Cell Biol, 2008. **84**: p. 843-870.

131. Warwicker, J. and H.C. Watson, *Calculation of the electric potential in the active site cleft due to alpha-helix dipoles.* J Mol Biol, 1982. **157**(4): p. 671-679.
132. Born, M., Z. Phys, 1920. **1**: p. 45.
133. Still, W.C., A. Tempczyk, R.C. Hawley, and T. Hendrickson, *Semianalytical treatment of solvation for molecular mechanics and dynamics.* J Am Chem Soc, 1990. **112**(16): p. 6127-6129.
134. Qiu, D., P.S. Shenkin, F.P. Hollinger, and W.C. Still, *The GB/SA Continuum Model for Solvation. A Fast Analytical Method for the Calculation of Approximate Born Radii.* J Phys Chem A, 1997. **101**(16): p. 3005-3014.
135. Chothia, C., *Hydrophobic bonding and accessible surface area in proteins.* Nature, 1974. **248**(446): p. 338-339.
136. Shrake, A., Rupley, J. A., *Environment and exposure to solvent of protein atoms. Lysozyme and insuline.* J Mol Biol, 1973. **79**(2): p. 351-364.
137. Gallicchio, E. and R.M. Levy, *AGBNP: an analytic implicit solvent model suitable for molecular dynamics simulations and high-resolution modeling.* J Comput Chem, 2004. **25**(4): p. 479-499.
138. Wagoner, J.A. and N.A. Baker, *Assessing implicit models for nonpolar mean solvation forces: the importance of dispersion and volume terms.* Proc Natl Acad Sci U S A, 2006. **103**(22): p. 8331-8336.
139. Berendsen, H.J.C., Postma, J.P.M., van Gunsteren, W.F., DiNola, A. Haak, J.R., *Molecular dynamics with coupling to an external bath.* J Chem Phys, 1984. **81**: p. 3684-3690.
140. Hoover, W.G., *Canonical dynamics: Equilibrium phase-space distributions.* Phys Rev A, 1985. **31**(3): p. 1695-1697.
141. Adelman, S.A., Doll, J.D., *Generalized Langevin equation approach for atom/solid-surface scattering: General formulation for classical scattering off harmonic solids.* J Chem Phys, 1976. **64**(6): p. 2375-2388.
142. Gilson, M.K., J.A. Given, B.L. Bush, and J.A. McCammon, *The statistical-thermodynamic basis for computation of binding affinities: a critical review.* Biophys J, 1997. **72**(3): p. 1047-1069.
143. Jarzynski, C., *Nonequilibrium Equality for Free Energy Differences.* Phys Rev Lett, 1997. **78**(14): p. 2690-2693.
144. Schwab, F., W.F. van Gunsteren, and B. Zagrovic, *Computational study of the mechanism and the relative free energies of binding of anticholesteremic inhibitors to squalene-hopene cyclase.* Biochemistry, 2008. **47**(9): p. 2945-2951.
145. Durrant, J.D. and J.A. McCammon, *Molecular dynamics simulations and drug discovery.* BMC Biol, 2011. **9**: p. 71.
146. Srinivasan, J., T.E. Cheatham, P. Cieplak, P.A. Kollman, and D.A. Case, *Continuum Solvent Studies of the Stability of DNA, RNA, and Phosphoramidate–DNA Helices.* J Am Chem Soc, 1998. **120**(37): p. 9401-9409.
147. Karplus, M. and J.N. Kushick, *Method for estimating the configurational entropy of macromolecules.* Macromolecules, 1981. **14**(2): p. 325-332.
148. Huang, N., C. Kalyanaraman, K. Bernacki, and M.P. Jacobson, *Molecular mechanics methods for predicting protein-ligand binding.* Phys Chem Chem Phys, 2006. **8**(44): p. 5166-5177.
149. Leach, A.R., Gillet, V.J, *Scoring Functions for Protein-Ligand Docking*, in *An Introduction to Cheminformatics.* 2003, Kluwer Academic Publishers. p. 174.

150. Verdonk, M.L., J.C. Cole, M.J. Hartshorn, C.W. Murray, and R.D. Taylor, *Improved protein-ligand docking using GOLD.* Proteins, 2003. **52**(4): p. 609-623.
151. Velec, H.G., Gohlke, Holger, H., Klebe, G., *DrugScore(CSD)-knowledge-based scoring function derived from small molecule crystal data with superior recognition rate of near-native ligand poses and better affinity prediction.* J Med Chem, 2005. **48**(20). p. 6296-6303
152. Huey, R., G.M. Morris, A.J. Olson, and D.S. Goodsell, *A semiempirical free energy force field with charge-based desolvation.* J Comput Chem, 2007. **28**(6): p. 1145-1152.
153. Pearlman, D.A. and P.S. Charifson, *Are Free Energy Calculations Useful in Practice? A Comparison with Rapid Scoring Functions for the p38 MAP Kinase Protein System.* J Med Chem, 2001. **44**(21): p. 3417-3423.
154. Trott, O. and A.J. Olson, *AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading.* J Comput Chem, 2010. **31**(2): p. 455-461.
155. Morris, G.M., R. Huey, W. Lindstrom, M.F. Sanner, R.K. Belew, D.S. Goodsell, and A.J. Olson, *AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility.* J Comput Chem, 2009. **30**(16): p. 2785-2791.
156. Sanner, M.F., *Python: a programming language for software integration and development.* J Mol Graph Model, 1999. **17**(1): p. 57-61.
157. Halgren, T.A., *Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94.* J Comput Chem, 1996. **17**(5-6): p. 490-519.
158. Wolf, L.K., *New software and Websites for the Chemical Enterprise.* Chem Eng News, 2009. **87**(31).
159. Bembenek, S.D., B.A. Tounge, and C.H. Reynolds, *Ligand efficiency and fragment-based drug discovery.* Drug Discov Today, 2009. **14**(5-6): p. 278-283.
160. D.A. Case, T.A.D., T.E. Cheatham, III, C.L. Simmerling, J. Wang, R.E. Duke, R. Luo, R.C. Walker, W. Zhang, K.M. Merz, B.P. Roberts, B. Wang, S. Hayik, A. Roitberg, G. Seabra, I. Kolossváry, K.F. Wong, F. Paesani, J. Vanicek, J. Liu, X. Wu, S.R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D.R. Roe, D.H. Mathews, M.G. Seetin, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko, and P.A. Kollman *AMBER 11 & AmberTools 1.5.* University of California, San Francisco, 2010.
161. Hornak, V., R. Abel, A. Okur, B. Strockbine, A. Roitberg, and C. Simmerling, *Comparison of multiple Amber force fields and development of improved protein backbone parameters.* Proteins, 2006. **65**(3): p. 712-725.
162. Wang, J., R.M. Wolf, J.W. Caldwell, P.A. Kollman, and D.A. Case, *Development and testing of a general amber force field.* J Comput Chem, 2004. **25**(9): p. 1157-1174.
163. Ryckaert, J.-P., G. Ciccotti, and H.J.C. Berendsen, *Numerical integration of the cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes.* J Comput Phys, 1977. **23**(3): p. 327-341.
164. Cheatham, T.E., B.R. Brooks, and P.A. Kollman, *Molecular Modeling of Nucleic Acid Structure: Electrostatics and Solvation*, in *Curr Protoc Nucleic Acid Chem*, 2001 Aug; Chapter 7: Unit 7.9
165. Weiser, J., P.S. Shenkin, and W.C. Still, *Approximate atomic surfaces from linear combinations of pairwise overlaps (LCPO).* J Comput Chem, 1999. **20**(2): p. 217-230.

166. Onufriev, A., D. Bashford, and D.A. Case, *Exploring protein native states and large-scale conformational changes with a modified generalized born model.* Proteins, 2004. **55**(2): p. 383-394.
167. Huo, S., I. Massova, and P.A. Kollman, *Computational alanine scanning of the 1:1 human growth hormone-receptor complex.* J Comput Chem, 2002. **23**(1): p. 15-27.
168. Gohlke, H., C. Kiel, and D.A. Case, *Insights into protein-protein binding by binding free energy calculation and free energy decomposition for the Ras-Raf and Ras-RalGDS complexes.* J Mol Biol, 2003. **330**(4): p. 891-913.
169. Roe, D.R. and T.E. Cheatham, *PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data.* J Chem Theory Comput, 2013.
170. Shao, J., S.W. Tanner, N. Thompson, and T.E. Cheatham, *Clustering Molecular Dynamics Trajectories: 1. Characterizing the Performance of Different Clustering Algorithms.* J Chem Theory Comput, 2007. **3**(6): p. 2312-2334.
171. Humphrey, W., A. Dalke, and K. Schulten, *VMD: visual molecular dynamics.* J Mol Graph, 1996. **14**(1): p. 33-38, 27-38.
172. Pettersen, E.F., T.D. Goddard, C.C. Huang, G.S. Couch, D.M. Greenblatt, E.C. Meng, and T.E. Ferrin, *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-1612.
173. Dundas, J., Z. Ouyang, J. Tseng, A. Binkowski, Y. Turpaz, and J. Liang, *CASTp: computed atlas of surface topography of proteins with structural and topographical mapping of functionally annotated residues.* Nucleic Acids Res, 2006. **34**(Web Server issue): p. W116-118.
174. Richards, F.M., *Areas, Volumes, Packing, and Protein Structure.* Annu Rev Biophys Bio, 1977. **6**: p. 151-176.
175. Connolly, M.L., *Analytical molecular surface calculation.* J Appl Crystallogr, 1977. **16**(5): p. 548-558.
176. Yaffe, E., D. Fishelovitch, H.J. Wolfson, D. Halperin, and R. Nussinov, *MolAxis: a server for identification of channels in macromolecules.* Nucleic Acids Res, 2008. **36**(Web Server issue): p. W210-215.
177. Gumbart, J., L.G. Trabuco, E. Schreiner, E. Villa, and K. Schulten, *Regulation of the protein-conducting channel by a bound ribosome.* Structure, 2009. **17**(11): p. 1453-1464.
178. Ghersi, D. and R. Sanchez, *Improving accuracy and efficiency of blind protein-ligand docking by focusing on predicted binding sites.* Proteins, 2009. **74**(2): p. 417-424.
179. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A.E. Mark, and H.J. Berendsen, *GROMACS: fast, flexible, and free.* J Comput Chem, 2005. **26**(16): p. 1701-1718.
180. Hernandez, M., D. Ghersi, and R. Sanchez, *SITEHOUND-web: a server for ligand binding site identification in protein structures.* Nucleic Acids Res, 2009. **37**(Web Server issue): p. W413-416.
181. Ando, T., S.-i. Inomata, and M. Yamamoto, *Lepidopteran Sex Pheromones The Chemistry of Pheromones and Other Semiochemicals I.* 2004, Top Curr Chem, 2004. **239**: p. 51-96.
182. Mori, K., *Pheromone synthesis.* Top Curr Chem, 2004. **239**: p. 1-50.
183. Sali, A. and T.L. Blundell, *Comparative protein modelling by satisfaction of spatial restraints.* J Mol Biol, 1993. **234**(3): p. 779-815.

184. Krissinel, E. and K. Henrick, *Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions.* Acta Crystallogr D Biol Crystallogr, 2004. **60**(Pt 12 Pt 1): p. 2256-2268.
185. Garcia-Sosa, A.T., C. Hetenyi, and U. Maran, *Drug efficiency indices for improvement of molecular docking scoring functions.* J Comput Chem, 2010. **31**(1): p. 174-184.
186. Kuntz, I.D., K. Chen, K.A. Sharp, and P.A. Kollman, *The maximal affinity of ligands.* Proc Natl Acad Sci U S A, 1999. **96**(18): p. 9997-10002.
187. He, X., G. Tzotzos, C. Woodcock, J.A. Pickett, T. Hooper, L.M. Field, and J.J. Zhou, *Binding of the general odorant binding protein of Bombyx mori BmorGOBP2 to the moth sex pheromone components.* J Chem Ecol, 2010. **36**(12): p. 1293-1305.
188. Hetenyi, C. and D. van der Spoel, *Blind docking of drug-sized compounds to proteins with up to a thousand residues.* FEBS Lett, 2006. **580**(5): p. 1447-1450.
189. Grater, F., B.L. de Groot, H. Jiang, and H. Grubmuller, *Ligand-release pathways in the pheromone-binding protein of Bombyx mori.* Structure, 2006. **14**(10): p. 1567-1576.
190. Rastelli, G., A. Del Rio, G. Degliesposti, and M. Sgobba, *Fast and accurate predictions of binding free energies using MM-PBSA and MM-GBSA.* J Comput Chem, 2010. **31**(4): p. 797-810.
191. Yang, T., J.C. Wu, C. Yan, Y. Wang, R. Luo, M.B. Gonzales, K.N. Dalby, and P. Ren, *Virtual screening using molecular simulations.* Proteins, 2011. **79**(6): p. 1940-1951.
192. Hou, T., J. Wang, Y. Li, and W. Wang, *Assessing the performance of the molecular mechanics/Poisson Boltzmann surface area and molecular mechanics/generalized Born surface area methods. II. The accuracy of ranking poses generated from docking.* J Comput Chem, 2011. **32**(5): p. 866-877.
193. Li, Z. and T. Lazaridis, *Water at biomolecular binding interfaces.* Phys Chem Chem Phys, 2007. **9**(5): p. 573-581.
194. Davrazou, F., E. Dong, E.J. Murphy, H.T. Johnson, and D.N. Jones, *New insights into the mechanism of odorant detection by the malaria-transmitting mosquito Anopheles gambiae.* J Biol Chem, 2011. **286**(39): p. 34175-34183.
195. Hallem, E.A. and J.R. Carlson, *Coding of odors by a receptor repertoire.* Cell, 2006. **125**(1): p. 143-160.
196. Wang, G., A.F. Carey, J.R. Carlson, and L.J. Zwiebel, *Molecular basis of odor coding in the malaria vector mosquito Anopheles gambiae.* Proc Natl Acad Sci U S A, 2010. **107**(9): p. 4418-4423.
197. Li, S., J.F. Picimbon, S. Ji, Y. Kan, Q. Chuanling, J.J. Zhou, and P. Pelosi, *Multiple functions of an odorant-binding protein in the mosquito Aedes aegypti.* Biochem Biophys Res Commun, 2008. **372**(3): p. 464-468.
198. Zhang, L. and J. Hermans, *Hydrophilicity of cavities in proteins.* Proteins, 1996. **24**(4): p. 433-438.
199. Vogt, R.G., R. Rybczynski, and M.R. Lerner, *Molecular cloning and sequencing of general odorant-binding proteins GOBP1 and GOBP2 from the tobacco hawk moth Manduca sexta: comparisons with other insect OBPs and their signal peptides.* J Neurosci, 1991. **11**(10): p. 2972-2984.
200. Qiao, H., X. He, D. Schymura, L. Ban, L. Field, F. Dani, E. Michelucci, B. Caputo, A. Torre, K. Iatrou, J.-J. Zhou, J.r. Krieger, and P. Pelosi,

*Cooperative interactions between odorant-binding proteins of Anopheles gambiae.* Cell Mol Life Sci, 2011. **68**(10): p. 1799-1813.
201. Krissinel, E. and K. Henrick, *Inference of macromolecular assemblies from crystalline state.* J Mol Biol, 2007. **372**(3): p. 774-797.
202. Steinbrecher, T., D.A. Case, and A. Labahn, *A multistep approach to structure-based drug design: studying ligand binding at the human neutrophil elastase.* J Med Chem, 2006. **49**(6): p. 1837-1844.
203. Gohlke, H. and D.A. Case, *Converging free energy estimates: MM-PB(GB)SA studies on the protein-protein complex Ras-Raf.* J Comput Chem, 2004. **25**(2): p. 238-250.
204. Homeyer, N. and H. Gohlke, *Free Energy Calculations by the Molecular Mechanics Poisson–Boltzmann Surface Area Method.* Mol Inf, 2012. **31**: p. 114-122.
205. Kongsted, J., P. Soderhjelm, and U. Ryde, *How accurate are continuum solvation models for drug-like molecules?* J Comput Aided Mol Des, 2009. **23**(7): p. 395-409.
206. Weis, A., K. Katebzadeh, P. Soderhjelm, I. Nilsson, and U. Ryde, *Ligand affinities predicted with the MM/PBSA method: dependence on the simulation method and the force field.* J Med Chem, 2006. **49**(22): p. 6596-6606.
207. Roelofs, W.L., W. Liu, G. Hao, H. Jiao, A.P. Rooney, and C.E. Linn, Jr., *Evolution of moth sex pheromones via ancestral genes.* Proc Natl Acad Sci U S A, 2002. **99**(21): p. 13621-13626.
208. Chintapalli, V.R., J. Wang, and J.A. Dow, *Using FlyAtlas to identify better Drosophila melanogaster models of human disease.* Nat Genet, 2007. **39**(6): p. 715-720.
209. Bikadi, Z. and E. Hazai, *Application of the PM6 semi-empirical method to modeling proteins enhances docking accuracy of AutoDock.* J Cheminform, 2009. **1**: p. 15.
210. Kaupp, U.B., *Olfactory signalling in vertebrates and insects: differences and commonalities.* Nat Rev Neurosci, **11**(3): p. 188-200.